Thu. Jun 6, 2019

Room B

International Session | International Session | [ES] E-2 Machine learning

[3B3-E-2] Machine learning: image recognition and generation

Chair: Masakazu Ishihata (NTT)
1:50 PM - 3:30 PM Room B (2F Main hall B)

[3B3-E-2-01] Design a Loss Function which Generates a Spatial configuration of Image In-betweening
○Paulino Cristovao[1], Hidemoto Nakada[1,2], Yusuke Tanimura[1,2], Hideki Asoh[2] (1. University of Tsukuba, 2. National Advanced Institute of Science and Technology of Japan (AIST))
1:50 PM - 2:10 PM

[3B3-E-2-02] One-shot Learning using Triplet Network with kNN classifier
○Mu Zhou[1,2], Yusuke Tanimura[2,1], Hidemoto Nakada[2,1] (1. University of Tsukuba, 2. Artifical Intelligence Research Center, National Institute of Advanced Institute of Technology)
2:10 PM - 2:30 PM

[3B3-E-2-03] Cycle Sketch GAN: Unpaired Sketch to Sketch Translation Based on Cycle GAN Algorithm
○Takeshi Kojima[1] (1. Peach Aviation Limited)
2:30 PM - 2:50 PM

[3B3-E-2-04] Conditional DCGAN's Challenge: Generating Handwritten Character Digit, Alphabet and Katakana
○Rina Komatsu[1], Tad Gonsalves[1] (1. Sophia University)
2:50 PM - 3:10 PM

[3B3-E-2-05] Sparse Damage Per-pixel Prognosis Indices via Semantic Segmentation
○Takato Yasuno[1] (1. Research Institute for Infrastructure Paradigm Shift (RIIPS))
3:10 PM - 3:30 PM

International Session | International Session | [ES] E-2 Machine learning

[3B4-E-2] Machine learning: social links

Chair: Lieu-Hen Chen (National Chi Nan University), Reviewer: Yasufumi Takama (Tokyo Metropolitan University)
3:50 PM - 5:30 PM Room B (2F Main hall B)

[3B4-E-2-01] Social Influence Prediction by a Community-based Convolutional Neural Network
Shao-Hsuan Tai[1], Hao-Shang Ma[1], ○Jen-Wei Huang[1] (1. National Cheng Kung University)
3:50 PM - 4:10 PM

[3B4-E-2-02] A Community Sensing Approach for User Identity Linkage
○Zexuan Wang[1], Teruaki Hayashi[1], Yukio Ohsawa[1] (1. Department of Systems Innovation, School of Engineering, The University of Tokyo)
4:10 PM - 4:30 PM

[3B4-E-2-03] Learning Sequential Behavior for Next-Item Prediction
○Na Lu[1], Yukio Ohsawa[1], Teruaki Hayashi[1] (1. The University of Tokyo)
4:30 PM - 4:50 PM

[3B4-E-2-04] Application of Unsupervised NMT Technique to Japanese--Chinese Machine Translation
○Yuting Zhao[1], Longtu Zhang[1], Mamoru Komachi[1] (1. Tokyo Metropolitan University)
4:50 PM - 5:10 PM

[3B4-E-2-05] Synthetic and Distribution Method of Japanese Synthesized Population for Real-Scale Social Simulations
○Tadahiko Murata[1], Takuya Harada[1] (1. Kansai University)
5:10 PM - 5:30 PM

Room H

International Session | International Session | [ES] E-3 Agents

[3H3-E-3] Agents: safe and cooperative society

Chair: Ahmed Moustafa (Nagoya Institute of Technology), Reviewer: Takayuki Ito (Nagoya Institute of Technology)
1:50 PM - 3:30 PM Room H (303+304 Small meeting rooms)

[3H3-E-3-01] An Autonomous Cooperative Randomization Approach to Prevent Attacks Based on Traffic Trends in the Communication Destination Anonymization Problem
○Keita Sugiyama[1], Naoki Fukuta[1] (1. Shizuoka University)
1:50 PM - 2:10 PM

[3H3-E-3-02] Cooperation Model for Improving Scalability of the Multi-Blockchains System
○Keyang Liu[1], Yukio Ohsawa[1], Teruaki Hayashi [1] (1. University of Tokyo, Graduate school of engineer)
2:10 PM - 2:30 PM

[3H3-E-3-03] Effect of Visible Meta-Rewards on Consumer Generated Media

○Fujio Toriumi[1], Hitoshi Yamamoto[2], Isamu Okada[3] (1. The University of Tokyo, 2. Rissho University, 3. Soka University)

2:30 PM - 2:50 PM

[3H3-E-3-04] Toward machine learning-based facilitation for online discussion in crowd-scale deliberation

○Chunsheng Yang[1], Takayuki Ito[2], Wen GU[2] (1. National Research Council Canada, 2. Nagoya Institute of Technology)

2:50 PM - 3:10 PM

[3H3-E-3-05] An automated privacy information detection approach for protecting individual online social network users

○Weihua Li[1], Jiaqi Wu[1], Quan Bai[2] (1. Auckland University of Technology, 2. University of Tasmania)

3:10 PM - 3:30 PM

Room J

International Session | International Session | [ES] E-4 Robots and real worlds

[3J3-E-4] Robots and real worlds: Human Interactions

Chair: Yihsin Ho (Takushoku University), Eri Sato-Shimokawara (Tokyo Metropolitan University)

1:50 PM - 3:10 PM Room J (201B Medium meeting room)

[3J3-E-4-01] Automatic Advertisement Copy Generation System from Images

○Koichi Yamagata[1], Masato Konno[1], Maki Sakamoto[1] (1. The University of Electro-Communications)

1:50 PM - 2:10 PM

[3J3-E-4-02] Eye-gaze in Social Robot Interactions

Koki Ijuin[2], ○Kristiina Jokinen Jokinen[1], Tsuneo Kato[2], Seiichi Yamamoto[2] (1. AIRC, AIST Tokyo Waterfront, 2. Doshisha University)

2:10 PM - 2:30 PM

[3J3-E-4-03] A Team Negotiation Strategy that Considers Team Interdependencies

○Daiki Setoguchi[1], Ahmed Moustafa[1], Takayuki Ito[1] (1. Nagoya Institute of Technology)

2:30 PM - 2:50 PM

[3J3-E-4-04] Identity Verification Using Face Recognition Improved by Managing Check-in Behavior of Event Attendees

○Akitoshi Okumura[1], Susumu Handa[1], Takamichi Hoshino[1], Naoki Tokunaga[1], Masami Kanda[1] (1. NEC Solution Innovators, Ltd.)

2:50 PM - 3:10 PM

International Session | International Session | [ES] E-2 Machine learning

[3B3-E-2] Machine learning: image recognition and generation

Chair: Masakazu Ishihata (NTT)

Thu. Jun 6, 2019 1:50 PM - 3:30 PM Room B (2F Main hall B)

[3B3-E-2-01] Design a Loss Function which Generates a Spatial configuration of Image In-betweening

○Paulino Cristovao[1], Hidemoto Nakada[1,2], Yusuke Tanimura[1,2], Hideki Asoh[2] （1. University of Tsukuba, 2. National Advanced Institute of Science and Technology of Japan (AIST)）

1:50 PM - 2:10 PM

[3B3-E-2-02] One-shot Learning using Triplet Network with kNN classifier

○Mu Zhou[1,2], Yusuke Tanimura[2,1], Hidemoto Nakada[2,1] （1. University of Tsukuba, 2. Artifical Intelligence Research Center, National Institute of Advanced Institute of Technology）

2:10 PM - 2:30 PM

[3B3-E-2-03] Cycle Sketch GAN: Unpaired Sketch to Sketch Translation Based on Cycle GAN Algorithm

○Takeshi Kojima[1] （1. Peach Aviation Limited）

2:30 PM - 2:50 PM

[3B3-E-2-04] Conditional DCGAN's Challenge: Generating Handwritten Character Digit, Alphabet and Katakana

○Rina Komatsu[1], Tad Gonsalves[1] （1. Sophia University）

2:50 PM - 3:10 PM

[3B3-E-2-05] Sparse Damage Per-pixel Prognosis Indices via Semantic Segmentation

○Takato Yasuno[1] （1. Research Institute for Infrastructure Paradigm Shift (RIIPS)）

3:10 PM - 3:30 PM

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Design a Loss Function which Generates a Spatial configuration of Image In-betweening

Paulino Crsitovao[*1] Hidemoto Nakada[*2*1] Yusuke Tanimura[*2*1] Hideki Asoh[*2]

[*1] University of Tsukuba
[*2] National Institute of Advanced Industrial Science and Technology of Japan

Instead of generating image inbetween directly from adjacent frames, we propose a method based on inbetweening in latent space. We design a simple loss function which generates a latent space that represent the spatial configuration of image inbetween. Contrary to the frame based methods, this model can make plausible assumption about the moving objects in the image and can capture what is not seen in the images. Our model has three networks, all based on variational autoencoder, sharing same weights. We validate this model on different synthetic datasets. We show the details of our network architecture and the evaluation results.

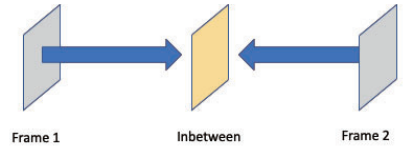
1. Introduction

For machines to become more intelligent and autonomous is essential that they understand the world around them, by being able to learn and understand the semantics present in the data. One way to approach this issue is by using generative models. These models can learn the patterns present in the data and generate new similar sample. This work seeks to discover latent representations present in data also design an objective function which generates the spatial configuration of image inbetween. Image inbetween attempts to generate image interpolation from nearby frames. The generated image has to preserve the spatial configuration of the moving objects. Up to now, optical flow [Yi 15],[Mémin 98] and convolutional neural networks [Amersfoort 17] have been proposed to generate image interpolation. Both methods generate image inbetween directly from adjacent frames 1. The result is blur images and loss of contextual information, also they cannot capture what is not present in the the frames. When generating image inbetween preserving the spatial location, shape, color is relevant for some application, for this reason we design a simple model that is able to preserve the contextual representations of objects between nearby frames. This model find scope in several areas such as movie and animation industries where they have to draw each individual frame and in image inpaiting.
In section 2 we describe our proposed model to generate image interpolation, in section 3, we show the results and we present conclusion in the last section.

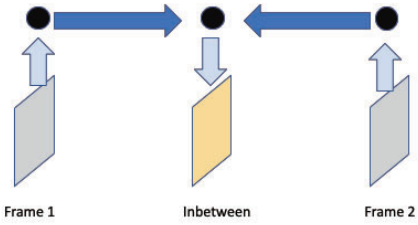
2. Proposed Method to Generate Image Inbetween

2.1 Model Overview

Our model is based on generative models, which have shown tremendous success in different field such as pattern recognition, image classification, natural language process and reinforcement learning. The proposed approach uses variational autoencoder to generate image inbetween.



(a) Existing Approach: Directly interpolate image from other frames



(b) Latent Model Interpolation

Figure 1: Comparison between existing approach and our

2.2 Proposed Loss Function to Generate Image Inbetween

Evaluating generative models means adjusting the internal weights of the network in order. to minimize an error measure. The error is usually given by a loss function. We optimize our network to minimize the distance between the image inbetween and ground truth in latent space. We follow a standard loss function for variational autoencoder. The model has three VAE each with its error function, we sum all errors function plus a an error function caused by the difference between the average latent space of the nearby frames and ground truth. We introduce a scalar hyper-parameter that we call coefficient α (below equation). The coefficient α is an adjustable parameter which express how much is relevant the difference between the Z1 and Z'. Next section we highlight the relevance of α.

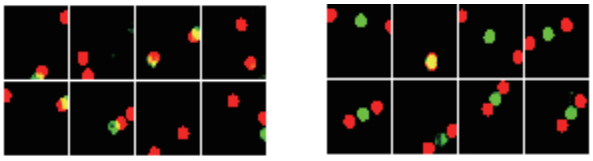
Contact: paulinocristovao86@gmail.com

$$l_{(X_0,X_1,X_2)} = l_{VAE}(X_0) + l_{VAE}(X_1) + l_{VAE}(X_2) + \alpha(D_{KL}(q_{(X_1)}||\frac{q_{(X_0)} + q_{(X_2)}}{2}))$$

2.3 Effects of Coefficient α

The adjustable hyper-parameter α modifies the traditional variational autoencoder objective function. It places a restriction on the latent space. This coefficient constraint the latent representations to generate a latent space which represent the spatial configuration of inbetween objects in the image. For $\alpha = 0$ represents the traditional VAE, no restriction is placed in the latent model, increasing the value of α means increasing restrictions on the latent representations.

When evaluating images, the motion of large objects seems easy to evaluate however, evaluating small motion is more complex. We aim to be able to detect small and large changes between frames.



(a) $\alpha = 0$,epochs=5000 (b) $\alpha = 100$,epochs=5000

Figure 2: Inbetween Image: Red dots: Nearby Images; Green dot: Inbetween Image

3. Experiments

3.1 Data Preprocessing

We used synthetic datasets. Two scenarios were tested: first where the object has one variable influencing its rotation which we name "one degree of freedom" and second having two variables "two degrees of freedom". The images were reshape into size of 32x32. For training and testing we randomly sample a triplet images by giving a certain interval among the frames.

3.2 Network Implementation

The base of our model follows a variational autoencoders (VAE), the network model has three VAEs 3, all sharing same weights to reduce the number of hyper-paremeters. The encoder has four convolutional layers, first layer (128 nodes), second(256 nodes), third (512 nodes), fourth (1024 nodes), kernel size = 4 and stride 2. The decoder has four deconvolutional layers, first layer (512 nodes), second (512 nodes), third (256 nodes), fourth (64 nodes) with same kernel size and stride. We input a triplet image. For this work we ignore the output of the nearby frames 3.

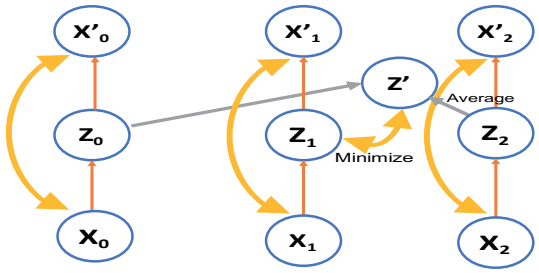


Figure 3: Network implementation

3.3 Reconstruction

The goal is to test the location accuracy

In this section, firstly we qualitatively demonstrate that our proposed model can reconstruct the input image. We tested the reconstruction object location, shape and color. Two scenarios is tested, on α equal to zero and α greater than zero.

3.3.1 One degree of freedom

Below results are for testing. We note that after strong coefficient $\alpha = 100$, the reconstruction test misses some features of the input data.

(a) $\alpha = 0$, epochs=10000 (b) $\alpha = 100$,epochs=10000

Figure 4: Teapot:1st row: Original Image, 2nd row:Reconstructed Image. Red box shows imperfect object reconstruction

3.3.2 Two degrees of freedom

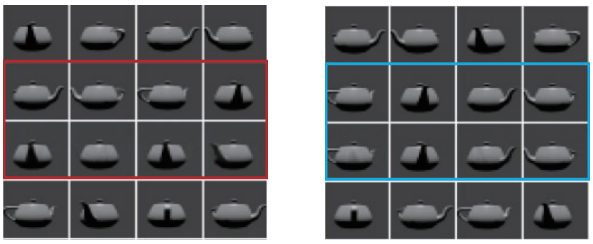
We increase the complexity of the data, the rotation of the object is influenced by two variables.

(a) $\alpha = 0$, epochs=10000 (b) $\alpha = 100$,epochs=10000

Figure 5: Teapot: 1st row:Original Image, 2nd row:Reconstructed Image

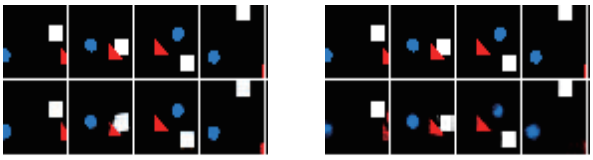
3.3.3 Multiple Objects

The task of reconstructing 3 objects seems complex for the model, since it has to capture the pattern of each object and make correspondent matching while interpolating.



(a) $\alpha = 0$,epochs=10000 (b) $\alpha = 100$,epochs=10000

Figure 7: Teapot-Testing: 1st row:first image, 2nd row: ground truth, 3rd row: Inbetween Image, 4th row: second image. The red square box shows that with $\alpha = 0$ we have imperfect inbetween, Blue box show the correct inbetween



(a) $\alpha = 0$, epochs=15000 (b) $\alpha = 100$,epochs=10000

Figure 6: Mulitple Objects: 1st row:Original Image, 2nd row:Reconstructed Image

3.4 Image Inbetween

As the model was able to reconstruct its input image, even with strong restriction placed in the latent space, next we qualitatively demonstrate the image inbetween generated by our model 7. Using two nearby images with large displacement from one image to other, with zero coefficient the image inbetween does not preserve the accurate spatial location of the object, increasing the coefficient the model is able to generate the perfect image inbetween as we will show in the next section.

3.4.1 One degree of freedom

We trained the framework with images or rotating on x-axis with one degree of freedom. The testing size is 360, the test images used in training and testing are distinct. The images generated by our approach $\alpha = 100$ presents a fair inbetween 7.

3.4.2 Two and Six degrees of freedom

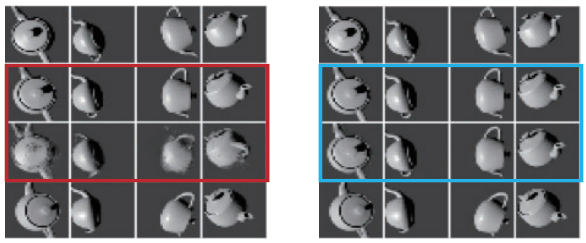
Previous examples we rotated the object in 360 degrees on x-axis, i.e. with one degree of freedom. It is easy to find the pattern of the data points as there are just 360 options or angles. We increased the complexity of the images by moving the object with two and six degrees of freedom. The results for two degrees are credible 8, while for six degrees (3 objects), the model does not perform well on testing phase 9.

3.5 Quantitative Evaluation

The goal here is to evaluate the complexity of the dataset in terms of its degree of freedom. We evaluate the same object in one degree and two degrees of freedom. The results indicates that the two degrees of freedom is more complex. Its MSE gives higher values 10.

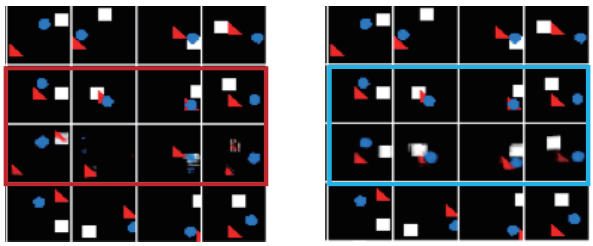
3.6 Linear Latent Space Interpolation

We sample pair of images x1 and x2 and project them into latent space z1 and z2 by sampling from the encoder,



(a) $\alpha = 0$,epochs=15000 (b) $\alpha = 100$,epochs=15000

Figure 8: 2D Teapot-Testing: 1st row:first image, 2nd row: ground truth, 3rd row: Inbetween Image, 4th row: second image. The red square box shows that with $\alpha = 0$ we have imperfect inbetween, Blue box show the correct inbetween

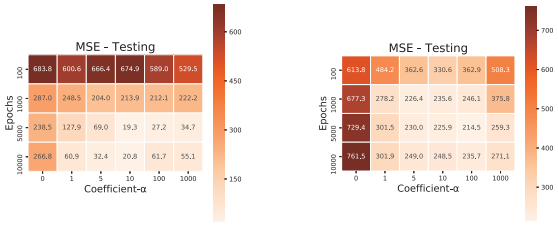


(a) $\alpha = 0$,epochs=15000 (b) $\alpha = 5$,epochs=15000

Figure 9: Multiple Objects - Testing: 1st row:first image, 2nd row: ground truth, 3rd row: Inbetween Image, 4th row: second image

then linearly interpolate between Z1 and Z2 and pass the intermediary points through the decoder to plot the input-space interpolations. The objective is to estimate the continuity in the latent space. Below figures show the generated smooth interpolation of two nearby points. The latent codes used to generate the nine intermediate images are equivalent to (P=0.9, to 0.1): We observe smooth transitions between pairs of examples, and intermediary images remain credible 11. This is an indicator that this model is not just restricting its probability mass exclusively around training examples, but rather has learned latent features that generalize well.

Linear latent space interpolation, indicate that there is a continuity in the latent space which allows a smooth interpolation. We show an example of 3 objects moving in random direction 12, we linearly interpolate the latent space and generate the possible trajectory between first frame and last frame. This model can predict a long-term frames and has the ability to capture their trajectory.



(a) one degree (b) two degrees

Figure 10: MSE loss

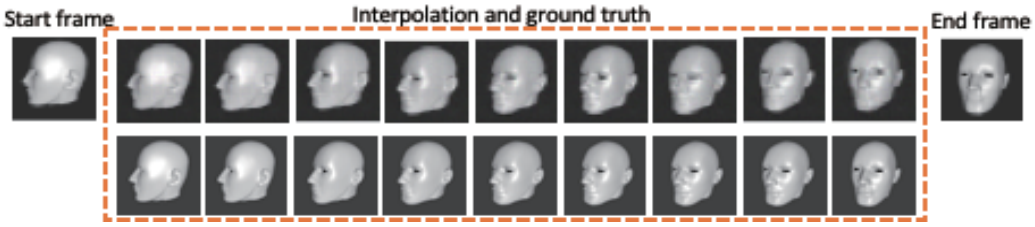


Figure 11: Linear Long-term Interpolation: Face turning to left side

Figure 12: Linear Long-term interpolation of 3 objects, we see the smooth interpolation from first frame and the last frame.

4. Related Work

The intention of unsupervised methods is to uncover the underlying latent representation of the data. Recent works on VAE [Higgins 17][Chen 18] [Berthelot 18] focus in disentangled the latent representations. This approach finds great application in scenario where there is a need to distinguish different characteristics present in the data, for instance, skin color, head pose, facial expression. A disentangled representation can be useful for natural tasks that require knowledge of the salient attributes of the data, which include tasks such as face and object recognition.
Our prosed model does not disentangle latent representations, it simply learn the pattern present in the data.

4.1 Improving Interpolation

While generating interpolation two fundamental characteristics have to be preserved: intermediate points along the interpolation are indistinguishable from real one and provide semantic and smooth morphing [Berthelot 18] The late characteristic is hard to achieve, for that reason [Berthelot 18] purpose a model based in variational autoencoder which introduce a regularizer which encourages interpolated data points to appear more indistinguishable from reconstructions of real data points. It is important to make a clear distinguish between image interpolation generated by latent model and our model. These latent model approaches cannot be used for our target application for the following reason, the dataset used by these models present some variation of the data, for instance in case of celebrity dataset mentioned earlier, it has many factors such as rotation of the head, skin color, age, gender, with or without glasses. The dataset we used does not present such characteristics in addition, we do not disentangle any specific factor of variations, we simply put a restriction on the latent model to generate an accurate image inbetween.

5. Conclusion

We present an alternative approach for generating an image inbetween by giving nearby frames which are non-consecutive images using a latent model. Our approach changes the Naive VAE objective function by introducing a hyper-parameter which constraint the latent representations. This model excels at predicting the image inbetwee in addition the model generalizes well for different datasets. For future, we will test this model on more complex data such as: Complex physical models, such as linked arms. Non-image data, for instance: text and audio data video i.e. video with fast motions and more moving objects. Finding better hyper-parameters between reconstruction and image inbetween.

Acknowledgement

This paper is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This work was supported by JSPS KAKENHI Grant Number JP16K00116.

References

[Amersfoort 17] Amersfoort, V., et al.: Frame Interpolation with Multi-Scale Deep Loss Functions and Generative Adversarial Networks, *arXiv preprint arXiv:1711.06045* (2017)

[Berthelot 18] Berthelot, D., et al.: Understanding and Improving Interpolation in Autoencoders via an Adversarial Regularizer, *arXiv preprint arXiv:1807.07543* (2018)

[Chen 18] Chen, T. Q., et al.: Isolating Sources of Disentanglement in Variational Autoencoders, *arXiv preprint arXiv:1802.04942* (2018)

[Higgins 17] Higgins, I., et al.: beta-vae: Learning basic visual concepts with a constrained variational framework, in *International Conference on Learning Representations* (2017)

[Mémin 98] Mémin, E. and Pérez, P.: Dense estimation and object-based segmentation of the optical flow with robust techniques, *IEEE Transactions on Image Processing*, Vol. 7, No. 5, pp. 703–719 (1998)

[Yi 15] Yi, C., Liyun, C., and Chunguang, L.: Moving Target Tracking Algorithm Based on Improved Optical Flow Technology, *Open Automation and Control Systems Journal*, Vol. 7, pp. 1387–1392 (2015)

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

One-shot Learning using Triplet Network with kNN classifier

Mu ZHOU[*1*2] Yusuke TANIMURA[*2*1] Hidemoto NAKADA[*2*1]

[*1]筑波大学
University of Tsukuba

[*2]産業技術総合研究所 人工知能研究センター
Artifical Intelligence Research Center, National Institute of Advanced Institute of Technology

We propose a triplet network with a kNN classifier for the problem of one-shot learning, in which we predict the query images by given single example of each class. Our triplet network learns a mapping from sample images to the Euclidean space. Then we apply kNN classifier on the embeddings generated by the triplet network to classify the query sample. Our method can improve the performance of one-shot classification with data augmentation by processing the images. Our experiments on different datasets which are based on MNIST dataset demonstrate that our approach provides a effective way for one-shot learning problems.

1. Introduction

Deep learning has shown great achievement in various tasks related to artificial intelligence such as object recognition [Girshick 15], image classification [Kaiming 15], and speech recognition [Yu 14]. However, huge amounts of labelled data is necessary for these deep neural network models to train on. In contrast, humans are capable of one-shot learning, which is to learn a concept from one or only a few training example, contrary to the normal practice of using a large amount of data. This is evident in the case of learning a new thing rapidly - humans have no problem recognizing the new category with one or a few direct observation. However, it is a challenging task for machine to solve the classification and recognition problem with very few labelled training data.

2. Related work

Several studies have investigated few-shot learning and one-shot learning, one special type neural network is Siamese Neworks [Koch 15]. The idea of the Siamese Network is based on distance metric learning which is to learn the distance metric from the input space of training data by a contrastive loss, then keep the samples belonging the same class close to each other and separate the dissimilar samples. The similar one is Triplet Network [Hoffer 15] which is composed of 3 parameter-shared convolutional neural networks.

Inspired by Siamese Networks and Triplet Networks, we improve the Triplet Network and use a triplet loss [Schroff 15] in our work. The loss function is to minimize the distance between the data with same label and maximize the distance between the data with different label. Before we get the embeddings trained on networks, we do data augmentation on the training set with only one sample. Then we make the prediction to the embedded query points by finding the nearest embedded support point by using k-Nearest Neighbor classifier. The procedure of the whole work is shown in Figure 1.

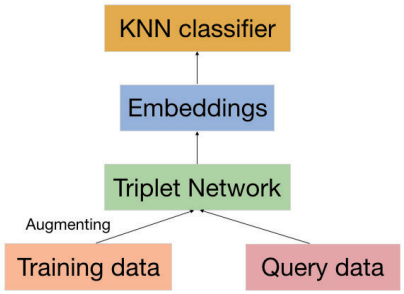


Figure 1: Prediction procedure.

3. Method

3.1 Triplet Network

In this research, we use the triplet network to learn the distance metric from inputs of triplet images. The triplet network is a horizontal concatenation triplet with 3 identical Convolutional Neural Networks (with shared parameters), these ConvNets are trained using triplets of inputs. The input triplet $(\vec{x_a}, \vec{x_p}, \vec{x_n})$ is composed of an anchor instance $\vec{x_a}$, a positive instance $\vec{x_p}$ (same class as the anchor), and a negative instance $\vec{x_n}$ (different class from the anchor). The network is then trained to learn an embedding function *f(x)* called triplet loss. The model architecture is shown in Figure 2.

3.1.1 Convolutional Networks

A series of breakthroughs in image classification came with the introduction of Convolutional Neural Networks (CNNs or ConvNets), where the image is input into a nested series of functions and convolved with filters, then output as feature vector. In our method, the ConvNet has 4 convolutional layers and is used as an embedding function. The ouput is passed through a fully connected layer resulting in a 128-dimensional embedding. In addition, we use ReLU as an activation function which is a common choice, especially for convolutional networks. The architecture of this ConvNet is as following:

- 1x{5x5-conv.layer (32 filters), 5x5-conv.layer (32 filters), batch normalization, max pool(2, 2), leaky relu,

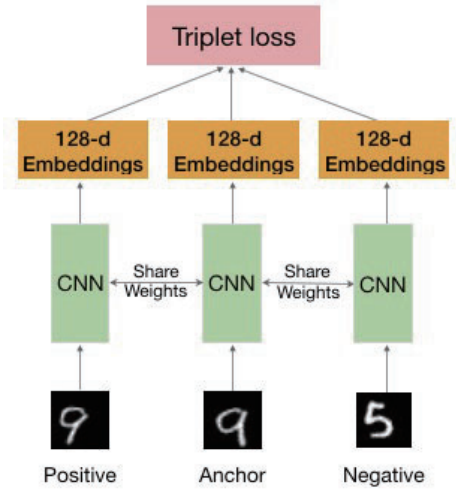


Figure 2: Triplet Network Model.

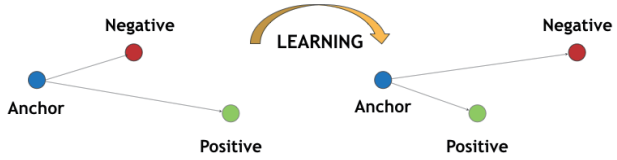


Figure 3: Triplet Loss Function.

dropout(0.25)},

- 1x{3x3-conv.layer (64 filters), 3x3-conv.layer (32 filters), batch normalization, max pool(2, 2), leaky relu, dropout(0.25)},
- 1x{fc-layer, batch normalization}.

3.1.2 Triplet Loss

Although we did not compare it to other loss function, we believe that the triplet loss is more suitable for this network, and triplet loss layer could improve the accuracy of ConvNets. A triplet loss is used to learn an embedding space for the images, such that embeddings of same class are close to each other, while those of different class are far away from each other. For the distance on the embedding space d, the loss of a triplet $(\vec{x_a}, \vec{x_p}, \vec{x_n})$ is:

$$L = max(d(x_a, x_p) - d(x_a, x_n) + \alpha, 0)$$

where α is a margin that is enforced between positive and negative pairs[Schroff 15]. In our research, the triplet loss minimizes the distance between the anchor and the positive, both of which have the same identity, and maximizes the distance between the anchor and a negative of a different identity, as shown in Figure 3.

3.2 kNN Classifier

The k-Nearest Neighbors algorithm is one of the simplest way to perform classification. Most kNN classifiers use Euclidean distances (also known as L2-norm distance) to measure the similarities between the instances which are represented as vector inputs. The L2-norm distance is as following:

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

In our research, after we trained the dataset (both train and test dataset) on the Triplet Network, we obtained the embeddings of the data, each of which is a 128-dimensional feature vector. Then we used PCA (Principal component analysis) to reduce the dimension of the feature vectors. Since these vector embeddings are represented in shared vector space, we can calculate the similarity between the vectors by using the vector distance. Finally we used kNN classifier to calculate the distance between the test point and all the training points by giving the feature vector of labelled training and unlabelled test data. We gained the best choice of k and choose the corresponding classification that appears most frequently as the predictive class.

3.3 Data Augmentation

Data augmentation is the most common solution for one-shot learning, since it can help to increase the amount of relevant data in the dataset and boost the performance of neural networks. In our research, we augmented the images in the training dataset. As a result, a large amount of training images was created, through different ways of processing or combination of multiple processing, such as random rotation, shifts and shear, etc.

4. Experiment

4.1 Dataset

MNIST database (Modified National Institute of Standards and Technology database) is a large database of handwritten digits that is commonly used for training various image processing systems. The MNIST database contains 60,000 images for training and 10,000 images for testing. Figure 4 presents some of the digits from MNIST dataset.

4.1.1 Initial Dataset

To setup the training dataset, we chose whole digit images with label 0 to 4, while we randomly selected simple digit image with the label 5 to 9 from the MNIST dataset. This initial dataset was used for our comparison experiment. The count of each label on initial training dataset is shown in Figure 5.

4.1.2 Augmented Dataset

In addition to the initial dataset, we generated another training dataset by the technique of data augmentation. In our experiment, we augmented the single image. Due to the limitation of some digit images, (i.e. digit 9 may be recognized as digit 6 after the 180-degree rotation,) we did the random rotation operation with only 30 degrees combined with random zoom and random shifts. To ensure similar appearance of the amount of each label, we enlarged the images several times with similar amount. The count of

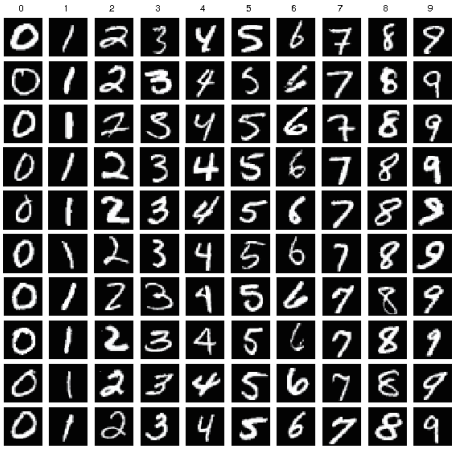


Figure 4: Samples from MNIST dataset.

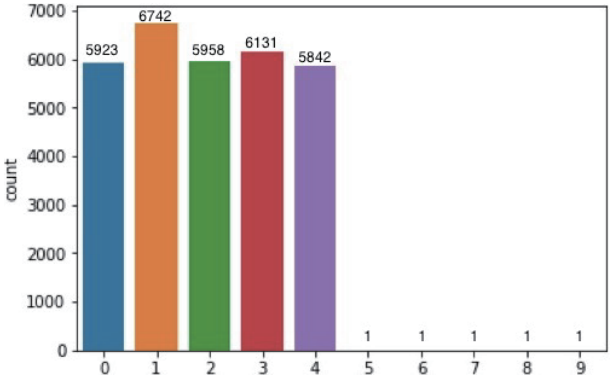


Figure 5: The initial dataset.

each label on augmented training dataset is shown in Figure 6.

4.2 Triplet Selection

Input triplets for Triplet Network were generated in two ways. One kind of triplets was produced by the augmented dataset, while another one was created by the initial dataset which was not augmented. For the first type, we randomly selected 1 sample (used as the anchor instance) from the dataset, then chose another one (used as the positive instance) from the same label. Then we randomly obtained the other sample (used as the negative instance) from any other label. Finally, we concatenated them as a triplet pair. However, for the other type created by initial dataset, we used the same image as the positive instance to overcome the limitation of lack of samples.

4.3 Results

We evaluated the performance of our model on above two datasets - initial dataset and augmented dataset, in order to judge the effectivity of data augmentation. To estimate the performance on Triplet Network in comparison to other model, we applied the CNN model on one-shot classification with the augmented dataset, as is mentioned above.

We obtained the embeddings of training points and test

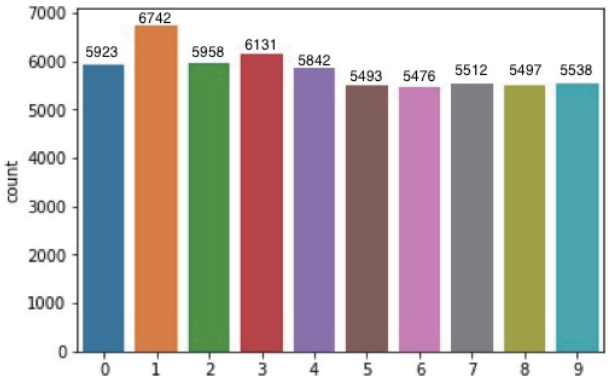


Figure 6: The augmented dataset.

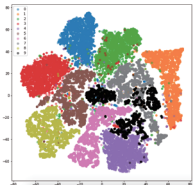


Figure 7: Embedding visualization of training points.

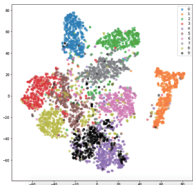


Figure 8: Embedding visualization of test points.

points from the Triplet Network, and we did visualization using t-SNE technique, as shown in Figure 7 and Figure 8. With these training points and test points, we evaluated the accuracy with different k ranging from 1 to 30, and selected the best choice of k. Figure 9 presents the accuracy of kNN classifier for different choice of k with augmented dataset in our Triplet Network model, and we get the best k (k=11) in this experiment. We predicted the label of test points with best k, and compared with the true label. The results are shown in Table 1, which present the accuracy of the test dataset with 1-shot classes (label 5 to label 9).

In our experiment on Triplet Network, the accuracy of the test dataset is 46.8% for 1-shot classes, while the accuracy

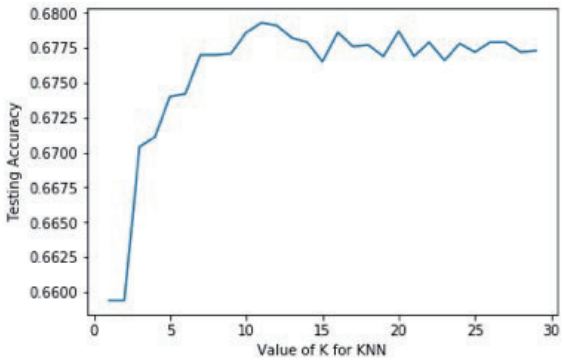


Figure 9: Accuracy of kNN classifer for different choices of k.

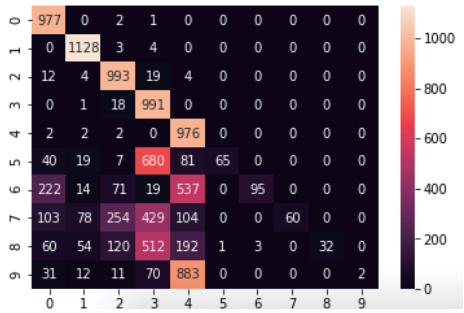


Figure 10: The result on TripletNN with initial dataset.

Figure 11: The result on TripletNN with augmented dataset.

is only 9.8% in the comparison experiment. This result suggests that data augmentation make sense and can obtain better prediction result than without data augmentation. In addition, the Triplet Network gives a better performance than the CNN model in this experiment.

Figure 10 and 11 show the results between actual labels and predicted labels in both datasets using Triplet Network. With regard to the accuracy of digit 9, it gets a low score since most of digit 9 are recognized as digit 4. This result implies that most written digit 9 are significantly similar to written digit 4, and the machine may not recognize them precisely with simple sample.

Method (dataset)	Accuracy					
	5	6	7	8	9	Average
TripletNN (not Agumented)	14%	18%	11%	6%	0%	9.8%
CNN (Augmented)	25%	26%	16%	24%	13%	20.8%
TripletNN (Augmented)	**42%**	**56%**	**66%**	**56%**	**14%**	**46.8%**

Table 1: Results of 1-shot classes.

5. Conclusion

In this work, we described how a Triplet Network model, inspired by the Siamese Network based on distance metric, can be used for one-shot learning. We used the embeddings of training points trained on kNN classifier and predict the label with the embedding of testing points by the classifier. We obtain significant improvement by the effectiveness of data augmentation. Of the 3 approaches tested, we achieved best results by augmenting the initial dataset with Triplet Network model. While in the contrast experiment on CNN model, data augmentation resulted accuracy of 20.8%. However, the experiment on Triplet model with initial dataset resulted accuracy of 9.8%, where almost all the data trained with 1 sample can not be recognized. This study therefore indicates that the benefits gained from data augmentation may work well on one-shot learning problem.

Although our experiment demonstrate a great improvement, the results are subject to certain limitations. For instance, since the differences between digit 9 and digit 4 are unable to be separated, most of digit 9 are recognized as digit 4 in the experiments. In addition, due to the computational constraint, our experiments were unable to explore how our approaches work on other much larger and complex datasets. Therefore, future work should focus on how to distinguish the difference between written digit 9 and digit 4 and how to enlarge the metric distance between digit 9 and 4. Furthermore, future studies need to be carried out in order to validate whether our approach does indeed help to solve the one-shot learning on other large and complex datasets, such as Fashion MNIST, Omniglot, Mini-Imagenet and e.t.

Acknowledgement

This paper is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This work was supported by JSPS KAKENHI Grant Number JP16K00116.

References

[Girshick 15] Girshick, R.: Fast R-CNN, *IEEE International Conference on Computer Vision (ICCV) 2015* (2015)

[Hoffer 15] Hoffer, E. and Ailon, N.: Deep metric learning using triplet network, *International Workshop on Similarity-Based Pattern Recognition* (2015)

[Kaiming 15] Kaiming, H., Xiangyu, Z., Shaoqing, R., and Jian., S.: Delving deep into rectifiers: Surpassing human- level performance on imagenet classification, *arXiv preprint arXiv:1502.01852* (2015)

[Koch 15] Koch, G., Zemel, R., and Salakhutdinov, R.: Siamese neural networks for one-shot image recognition, *ICML Deep Learning Workshop* (2015)

[Schroff 15] Schroff, F., Kalenichenko, D., and Philbin., J.: Facenet: A unified embedding for face recognition and clustering, *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015)

[Yu 14] Yu, D. and Deng, L.: *Automatic Speech Recognition: A Deep Learning Approach*, Springer Publishing Company (2014)

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Cycle Sketch GAN: Unpaired Sketch to Sketch Translation Based on Cycle GAN Algorithm

Takeshi Kojima

Peach Aviation Limited

Unlike pixel image generation, sketch drawing generative model outputs a sequence of pen stroke information. This paper proposes Cycle Sketch GAN: the first model that learns to translate a sketch drawing from source domain to target domain in the absence of paired dataset. Based on Cycle GAN algorithm, this model uses Transformer Encoder architecture in generators. Transformer Encoder feeds the input stroke information in source domain and generates the parameters for output distribution, from which the stroke information in target domain is sampled by reparameterization trick. The negative log likelihood of the distribution is used as cycle consistency loss. This model is trained and evaluated by some QuickDraw datasets. Qualitative evaluation shows that this model can practically translate sketch drawings from source domain to target domain. Quantitative evaluation by user study showed that 42 % of the translated sketches is recognizable compared to 71 % of the human sketches.

1. Introduction

Unlike pixel image generation, sketch drawing generative model outputs a sequence of pen stroke information (See section 2.1 for details). Some recent researches focused on creating sketch drawing generative model [Ha 18][Song 18] by neural network. However, the research of sketch to sketch translation, which aims to transform a sketch from source domain to target domain especially without supervised dataset, was not yet conducted.

This paper proposes Cycle Sketch GAN: the first model that learns to translate a sketch drawing from source domain to target domain in the absence of paired dataset. Specifically, this unsupervised learning model can change a sketch drawing's partial shapes characteristic to source domain into ones characteristic to target domain while keeping the common features unchanged (See an example in Fig.1). To train this model, we need to prepare 2 domain datasets, but each data in one domain does not need to have paired data in the other domain. This model is based on Cycle GAN algorithm[Zhu 17]. However, several changes are implemented to solve the following 2 problems specific to sketch drawing process.

The first problem of unsupervised sketch to sketch translation is that the sketch drawing is a process of generating a sequence of stroke vector representations. Therefore, Convolutional Neural Network(CNN), which is used in basic Cycle GAN, should not be used for this solution. Cycle Sketch GAN solves this problem by using Transformer[Vaswani 17] Encoder architecture in a generator. Transformer has self-attention mechanisms and recently succeeded in improvements of several sequential data processing tasks mainly in NLP. Note that Transformder Decoder is not used in this paper because we have no supervised data. Instead, Transformer Encoder feeds input and directly generates sequential outputs.

The second problem is that drawing sketch requires the accurate generation of strokes. Therefore, cycle consistency loss function has to be like the form of reconstruction loss as in [Ha 18], instead of L1 or L2 Norm. Cycle Sketch GAN solves this problem as follows: Transformer Encoder feeds the input stroke in source domain and generates the parameters for output distributions. Stroke information in target domain is sampled from the distribution by reparameterization trick to enable backpropagation for the training. The negative log likelihood of the distribution is used as cycle consistency loss to improve accuracy.

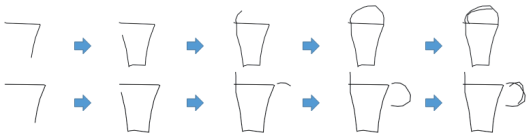


Figure 1: A visualized example of sequential stroke output process by Cycle Sketch GAN. This model can learn to translate a sketch drawing from source domain (Top: bucket) to target domain (Bottom: cup) in the absence of paired datasets. The common feature between 2 drawings (the body of bucket and cup) is unchanged.

2. Methods

2.1 Data Format

Based on [Ha 18], the data format of a sketch drawing for this model is a sequence of pen stroke actions $s = (s1, ...s_i, ..., s_{N_{max_s}})$, where $s_i = (\Delta_{x_i}, \Delta_{y_i}, p1_i, p2_i, p3_i)$. $\Delta_{x_i}, \Delta_{y_i}$ are the offset distance of ith pen movement in the direction of x axis and y axis. $p1_i, p2_i, p3_i$ are binary one-hot vector of 3 possible states at ith movement[*1]. $p1_i$ is an indicator that the pen is touching the paper for the ith pen movement. $p2_i$ is an indicator that the pen is lifted from the paper for the ith pen movement. $p3_i$ is an indicator that the drawing has ended. In case of $p3_i = 1$, Δ_{x_i} and Δ_{y_i} are defined to be 0.

2.2 Generator Architecture

This model uses Transformer[Vaswani 17] Encoder architecture as a generator to translate a sketch drawing stroke representation of domain S_A to domain S_B, and vice versa. Specifically, $s_A \in S_A$ is fed into one generator and translated to $\tilde{s_B} \in S_B$. In the same way, $s_B \in S_B$ is fed into the other generator and translated to $\tilde{s_A} \in S_A$. The architectures of these 2 generators are the same. This section omits the subscript A and B for simplicity.

Contact: t.kojima@flypeach.com

*1 [Ha 18] defines $p1_i, p2_i, p3_i$ as binary one-hot vector of 3 possible states at $i + 1$th movement.

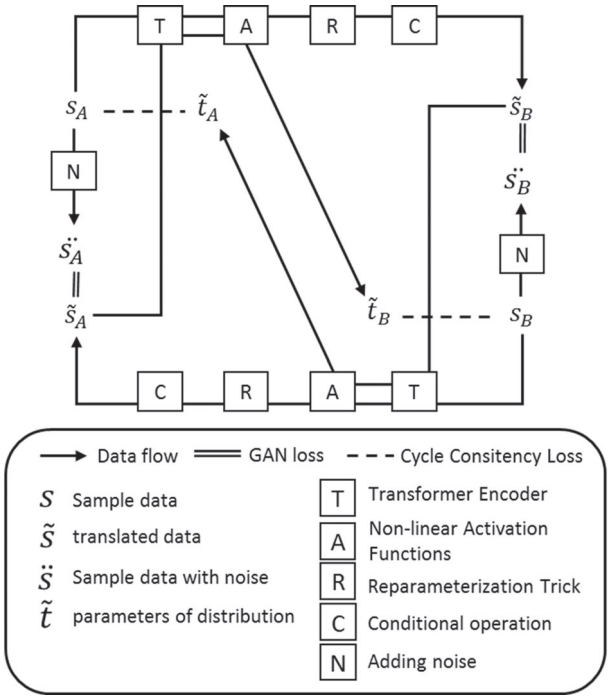


Figure 2: Overall image of Cycle Sketch GAN

The input s is concatenated with Positional Encoding [Vaswani 17] whose dimension size for ith position is ${N_{PE}}^{*2}$. The concatenated representation is fed into Transformer Encoder as input. Transformer Encoder has N_{TL}multiple stacked layers, each of which consists of N_{TA} multi-head attentions and a feed-forward network with dimension size of $4N_{TA}$, followed by a position-wise linear projection for the output below.

$$\begin{aligned}&(\mu_x, \mu_y, \hat{\sigma_x}, \hat{\sigma_y}, \hat{\rho_{xy}}, q1, q2, q3)_1 \\ &\quad .... \ (\mu_x, \mu_y, \hat{\sigma_x}, \hat{\sigma_y}, \hat{\rho_{xy}}, q1, q2, q3)_{N_{max_s}} \\ &\quad = TransformerEncoder([s; PE]) \end{aligned} \quad (1)$$

As described in [Ha 18], $(\mu_x, \mu_y, \hat{\sigma_x}, \hat{\sigma_y}, \hat{\rho_{xy}})$ are defined as the parameters for a bivariate normal distribution to describe Δx and Δy. $(q1, q2, q3)$ are the categorical distribution parameters to model the ground truth data $(p1, p2, p3)$. The following nonlinear functions are required to ensure that standard deviations are non-negative, that the correlation values are limited between -1 and 1.

$$\sigma_x = exp(\hat{\sigma_x}), \ \sigma_y = exp(\hat{\sigma_y}), \ \rho_{xy} = tanh(\hat{\rho_{xy}}) \quad (2)$$

The reparameterization trick for bivariate normal distribution is applied using $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho_{xy})$ to sample single value for each ith pen movements.

$$\begin{aligned}\begin{bmatrix} \tilde{\Delta x_i} \\ \tilde{\Delta y_i} \end{bmatrix} &= \begin{bmatrix} \mu_{x_i} \\ \mu_{y_i} \end{bmatrix} + L_i \begin{bmatrix} \epsilon_{x_i} \\ \epsilon_{y_i} \end{bmatrix} \\ where \ \ L_i &= \begin{bmatrix} \sigma_{x_i} & 0 \\ \rho_{xy_i}\sigma_{y_i} & \sqrt{1-{\rho_{xy_i}}^2}\sigma_{y_i} \end{bmatrix} \\ \epsilon_{x_i}, \epsilon_{y_i} &\sim N(0,1), \ \epsilon_{x_i}, \epsilon_{y_i} \in \mathbb{R} \end{aligned} \quad (3)$$

L_i is a lower triangular matrix after cholesky decomposition of covariance matrix of bivariate normal distribution.

∗2 Considering the case $N_{PE} > 5$, we use concatenation instead of addition.

Categorical reparameterization for $(q1, q2, q3)_i = q_i$ is also applied by using Gumbel Softmax[Jang 17] with temperature τ.

$$\begin{aligned}\tilde{q_i} &= softmax((q_i + g_i)/\tau) \\ where \ \ g_i &= -\log(-\log(u_i)) \\ u_i &\sim Uniform(0,1), \ u_i \in \mathbb{R}^3 \end{aligned} \quad (4)$$

In order to deceive discriminators as much as possible, generators calculate the following position-wise conditional operation before the data is fed into Discriminator.

$$\tilde{s_i} = \begin{cases} (0, 0, \tilde{q1}, \tilde{q2}, \tilde{q3})_i & if \ max(\tilde{q1}_i, \tilde{q2}_i, \tilde{q3}_i) = \tilde{q3}_i \\ (\tilde{\Delta x}, \tilde{\Delta y}, \tilde{q1}, \tilde{q2}, \tilde{q3})_i & otherwise \end{cases} \quad (5)$$

Here, for simplification, the sequence of functions (1) and (2) can be defined as $t = (t1, ..., t_i, ..., t_{N_{max_s}}) = G(s)$, where $t_i = (\mu_x, \mu_y, \sigma_x, \sigma_y, \rho_{xy}, q1, q2, q3)_i$. The sequence of functions (3), (4) and (5) can also be defined as $\tilde{s} = (\tilde{s}_1, \tilde{s}_2, ..., \tilde{s}_i, ..., \tilde{s}_{N_{max_s}}) = H(t)$. By using these expressions, 2 generators can be defined as:

$$\tilde{s_B} = H_B(G_B(s_A)), \ \ \tilde{s_A} = H_A(G_A(s_B)) \quad (6)$$

Note that the function $H_A(\cdot)$, $H_B(\cdot)$ does not contain any neural networks to be optimized.

2.3 Discriminator Architecture

The discriminator D is a multilayer convolutional neural network with instance normalization. Specifically, the discriminator regards the input s or $\tilde{s}$ as an image with height=N_{max_s}, width=5 and depth=1. The jth layer of D convolves the input of the layer with kernel size $(N_{Dk-height_j}, \ N_{Dk-weight_j})$ and stride size = N_{Ds_j} without padding, and outputs the tensor with channel size N_{Dc_j}, followed by instance normalization and activation. As an exception, the first layer and final layer does not have instance normalization. The final layer also has no activation function due to the restriction of the architecture of the discriminator in improved Wasserstein GAN optimization process [Gulrajani 17]. 2 discriminators, D_A for data A and D_B for data B are implemented with the same architecture described above.

2.4 Objective function

The objective function of this model contains adversarial losses and cycle consistency losses for both domains[Zhu 17]. As for the adversarial loss, this model uses improved Wasserstein GAN [Gulrajani 17] instead of the normal GAN [Goodfellow 14] to avoid mode collapse and to stabilize the training. The adversarial loss for data A is:

$$\begin{aligned}&\mathcal{L}^A_{GAN}(G_A, D_A, S_B, S_A) \\ &\quad = \mathbb{E}_{s_B}[D_A(H_A(G_A(s_B)))] - \mathbb{E}_{\ddot{s}_A}[D_A(\ddot{s}_A)] \\ &\qquad + \textit{gradient penalty for WGAN-GP} \\ &where \ \ddot{s}_A = noise(s_A) \end{aligned} \quad (7)$$

$noise(\cdot)$ is a function that adds random noises $U(-0.01, 0.01)$ into $(p1, p2, p3)$, and also into $(\Delta_{x_i}, \Delta_{y_i})$ if $q3_i = 1$ to prevent D from concentrating too much on discrete variables. The same as true for the adversarial loss of data B, $\mathcal{L}^B_{GAN}(G_B, D_B, S_A, S_B)$.

As for cycle consistency loss of data A, firstly the following cycled parameters are introduced.

$$(\tilde{t}_{A,1}, ..., \tilde{t}_{A,i}, ...\tilde{t}_{A,N_{max_s}}) = G_A(H_B(G_B(s_A)))$$
$$where\ \tilde{t}_{A,i} = (\tilde{\mu}_x^A, \tilde{\mu}_y^A, \tilde{\sigma}_x^A, \tilde{\sigma}_y^A, \tilde{\rho}_{xy}^A, \tilde{q1}^A, \tilde{q2}^A, \tilde{q3}^A)_i \quad (8)$$

Then, cycle consistency loss for data A is expressed as follows:

$$\mathcal{L}_{cyc}^A(G_A, G_B, S_A)$$
$$= \mathbb{E}_{s_A}\Big[- \frac{1}{N_{\max_s}} \sum_{i=1}^{N_s} \log\Big(\mathcal{N}(\Delta x_i^A, \Delta y_i^A \mid \tilde{w}_i^A)\Big)$$
$$- \frac{1}{N_{\max_s}} \sum_{i=1}^{N_{\max_s}} \sum_{k=1}^{3} pk_i^A \log(qk_i^{'A})\Big]$$
$$where\ \tilde{w}_i^A = (\tilde{\mu}_x^A, \tilde{\mu}_y^A, \tilde{\sigma}_x^A, \tilde{\sigma}_y^A, \tilde{\rho}_{xy}^A)_i$$
$$q1_i^{'A}, q2_i^{'A}, q3_i^{'A} = softmax(\tilde{q1}_i^A, \tilde{q2}_i^A, \tilde{q3}_i^A) \quad (9)$$

$\mathcal{N}(\Delta x_i^A, \Delta y_i^A \mid \tilde{w}_i^A)$ is the probability distribution function for a bivariate normal distribution. N_s is the point of last stroke in the sketch. This cycle consistency loss function is the same as the reconstruction loss function of [Ha 18] except that the distribution of $(\Delta x, \Delta y)$ is not modeled as a Gaussian mixture model (GMM). It can be said that the function is a special case of GMM size = 1. The same as true for the cycle consistency loss for data B, $\mathcal{L}_{cyc}^B(G_B, G_A, S_B)$.

The final objective function is:

$$\begin{aligned}\mathcal{L}(G_A, G_B, D_A, D_B) &= \mathcal{L}_{GAN}^A(G_A, D_A, S_B, S_A) \\ &+ \mathcal{L}_{GAN}^B(G_B, D_B, S_A, S_B) \\ &+ \lambda\mathcal{L}_{cyc}^A(G_A, G_B, S_A) \\ &+ \lambda\mathcal{L}_{cyc}^B(G_B, G_A, S_B) \quad (10)\end{aligned}$$

We aim to solve:

$$G_A^*, G_B^* = \arg\min_{G_A, G_B} \max_{D_A, D_B} \mathcal{L}(G_A, G_B, D_A, D_B) \quad (11)$$

3. Experimtents

3.1 Dataset

To evaluate Cycle Sketch GAN, full size of "Sketch-RNN QuickDraw Dataset" is used. QuickDraw Dataset contains hundreds of classes of sketch drawings. Each class is a dataset of more than 70K training samples and 2.5K test samples. In this paper, the following 6 classes are picked up from QuickDraw and made pairs: (bucket, cup), (suitcase, envelope), (sock, rollerskates). Note that each data in one class does not have paired data in the other class. The data format is changed according to section 2.1.

This experiment only uses the data whose size of the stroke actions does not exceed $N_{max_s} = 50$. Moreover, in order to equalize the training dataset size between paired classes, training data were randomly sampled from the class whose dataset size is larger than the other.

3.2 Implementation details

As for generators, the dimension size of Positional Encoding is set to be $N_{PE} = 251$. Transformer layer size is $N_{TL} = 12$, and the multihead attention size is $N_{TA} = 4$. As for the Discriminator, the hyper parameter values are set as followed:
$(N_{Dk-height_1}, N_{Dk-weight_1}, N_{Ds_1}, N_{Dc_1}) = (5, 5, 1, 128)$
$(N_{Dk-height_2}, N_{Dk-weight_2}, N_{Ds_2}, N_{Dc_2}) = (10, 1, 2, 256)$
$(N_{Dk-height_3}, N_{Dk-weight_3}, N_{Ds_3}, N_{Dc_3}) = (10, 1, 2, 512)$
$(N_{Dk-height_4}, N_{Dk-weight_4}, N_{Ds_4}, N_{Dc_4}) = (5, 1, 1, 1)$.

Figure 3: The original sketches by human (Row A, B) and the translated sketches by Cycle Sketch GAN (Row B', A').

All the activation functions in generators and discriminators are Leaky-ReLU activation. The temperature τ of Gumbel Softmax is set to be 1. λ in the final objective function is also set to be 1.

While training, minibatch SGD is used and the Adam optimizer is applied for the optimization process with $\beta_1 = 0$ and $\beta_2 = 0.9$ [Gulrajani 17]. The learning rate is fixed with 0.00005 during the training. The minibatch size is 128 and the iteration size is 30000 for generators. Discriminators and generators are mutually trained with the number of iterations of Discriminators per iteration of generators be set 5[Gulrajani 17].

Before generators feed data, the offsets $(\Delta x, \Delta y)$ in each classes is normalized using a single scaling factor [Ha 18] to adjust the offsets in the training set to have a standard deviation of 1. The test dataset is also normalized by that single scaling factor which is calculated by the training set of the same class.

2 types of data augmentation are applied into the training data for every iteration[Ha 18]. The first one is to stretch Δx and Δy respectively by multiplying random value drawn from $U(0.85, 1.15)$. The second one is to drop out strokes by dropping each point within line segments with a probability of 0.1.

Residual Dropout[Vaswani 17] is applied into Transformer Encoder with dropout rate = 0.1 during the training. At inference time, the dropout rate = 0, and also $\epsilon_x, \epsilon_y, g = 0$ in the reparameterization trick to produce the optimal translated drawings $\tilde{s_B}^* = H_B(G_B^*(s_A))$ and $\tilde{s_A}^* = H_A(G_A^*(s_B))$.

3.3 Results

3.3.1 Qualitative Evaluation

3 Cycle Sketch GAN models are trained by using 3 dataset pairs in QuickDraw (See section 3.1). Fig. 3 shows some examples of the translated sketch drawings from test data by the trained models. A and B rows are the randomly chosen sample drawings from each class test datasets. The drawings in B' rows are respectively the translated sketch drawings from the above A drawings. The

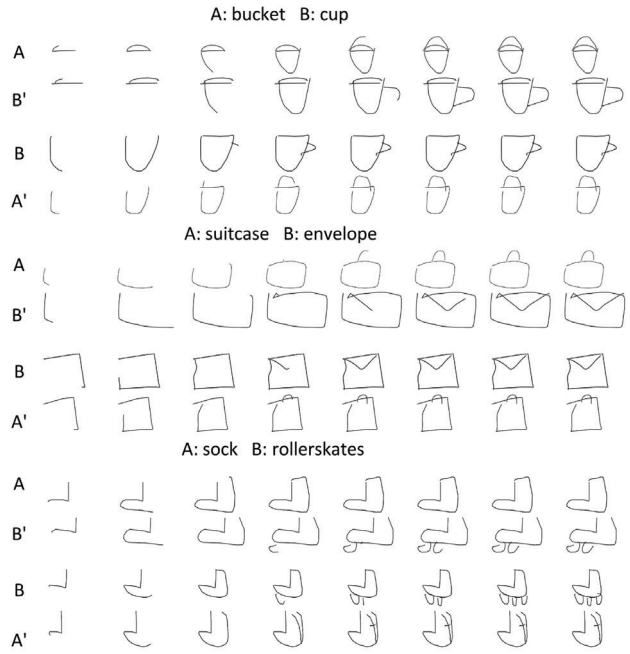


Figure 4: The stroke order of an original sketch by human (A, B) and the translated sketch by Cycle Sketch GAN (B', A').

same is true for A' and B. It shows that the common features are unchanged between the test sample drawing and the translated drawing, but the partial shapes characteristic to source domain are successfully transformed into ones characteristic to target domain, although some samples are failed to be translated.

Fig. 4 compares the order of the strokes between some test sample data and the translated data. Each row describes the snapshots of the strokes at the point of $i = 6, 12, 18, 24, 30, 36, 42, 48$ from the left to right. It clearly shows that the stroke order between original sketch and translated sketch are very similar from the start until some point. After that point, the stroke becomes different to express the characteristic feature for each domains.

3.3.2 Quantitative Evaluation

Class	Cycle Sketch GAN	Human
bucket	32.8($\pm$30.5)	57.6($\pm$25.9)
cup	43.0($\pm$28.5)	75.2($\pm$31.3)
suitcase	49.0($\pm$30.3)	54.4($\pm$27.0)
envelope	63.1($\pm$23.1)	85.8($\pm$18.7)
sock	38.4($\pm$26.1)	71.8($\pm$23.9)
rollerskates	30.8($\pm$14.8)	82.0($\pm$14.7)
Average	42.9	71.1

Table 1: User survey result

User study was also conducted on Amazon Mechanical Turk (AMT) to test the quality of translated sketch drawings. For each class, the survey results were collected from 20 participants. Specifically, participants see drawings one by one, and were asked "Do you think the drawing is (class name) ? ". Participants clicked on "yes" or "no" for the answer. For each dataset class, 55 drawings are surveyed, which consists of 25 of real sample sketches by humans, 25 of translated sketches by Cycle Sketch GAN, and 5 of trials (apparently irrelevant class sketches to check whether the participant's response was reliable or not). The order of showing these drawings are randomized.

Table. 1 shows the survey result, which lists the average percentage (and the standard deviation) of answering "yes" for each class surveys. Overall, 42 % of the translated sketches is recognizable compared to 71 % of the human sketch. There is apparent that some easy sketches such as envelope got higher recognition rate, while some complex sketches such as rollerskates got lower rate.

4. Conclusion and Future Work

This paper proposed Cycle Sketch GAN, the first model that learns to translate a sketch drawing from source domain to target domain in the absence of paired dataset. Qualitative and quantitative evaluation by some QuickDraw datasets demonstrates the effectiveness of this model. As a future work, the model needs to be improved to be able to translate complicated sketch drawings with higher quality. Using GMM as output distributions, or using Encoder-Decoder architecture as generators such as Seq2Seq might be effective. Furthermore, there might be a possibility that this unsupervised learning approach can be applied into other tasks that requires sequential data generation, such as unsupervised language translation[Lample 18], even though we need quite a lot of model changes and improvements.

References

[Goodfellow 14] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: Generative Adversarial Nets, in *Advances in Neural Information Processing Systems 27*, pp. 2672–2680 (2014)

[Gulrajani 17] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C.: Improved Training of Wasserstein GANs, in *Advances in Neural Information Processing Systems 30*, pp. 5767–5777 (2017)

[Ha 18] Ha, D. and Eck, D.: A Neural Representation of Sketch Drawings, in *International Conference on Learning Representations* (2018)

[Jang 17] Jang, E., Gu, S., and Poole, B.: Categorical Reparameterization with Gumbel-Softmax (2017)

[Lample 18] Lample, G., Ott, M., Conneau, A., Denoyer, L., and Ranzato, M.: Phrase-Based & Neural Unsupervised Machine Translation, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 5039–5049 (2018)

[Song 18] Song, J., Pang, K., Song, Y., Xiang, T., and Hospedales, T. M.: Learning to Sketch With Shortcut Cycle Consistency, in *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018*, pp. 801–810 (2018)

[Vaswani 17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I.: Attention is All you Need, in *Advances in Neural Information Processing Systems 30*, pp. 5998–6008 (2017)

[Zhu 17] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, in *Computer Vision (ICCV), 2017 IEEE International Conference on* (2017)

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Conditional DCGAN's Challenge: Generating Handwritten Character Digit, Alphabet and Katakana

Rina Komatsu[*1] Tad Gonsalves[*1]

[*1] Sophia University

Developing deep learning models has a great potential in assisting human tasks involving design and creativity. This study deals with generating handwritten characters using deep learning techniques. The task is not simply generating images randomly, but generating them conditionally, making a distinction according to the UI designates. To solve this task, we constructed the Conditional DCGAN model which includes the techniques from DCGAN and Conditional GAN. We tried training the models to be able to generate conditional images by adding label information as input to the Generator. Deep learning experiments were performed using 141319 training data consisting of 96 kinds of characters including digits, Roman alphabets and Katakana. The Generator trained by inputting random noise concatenated with the 96 kinds of characters, could generate each kind of character by just adding the appropriate label information.

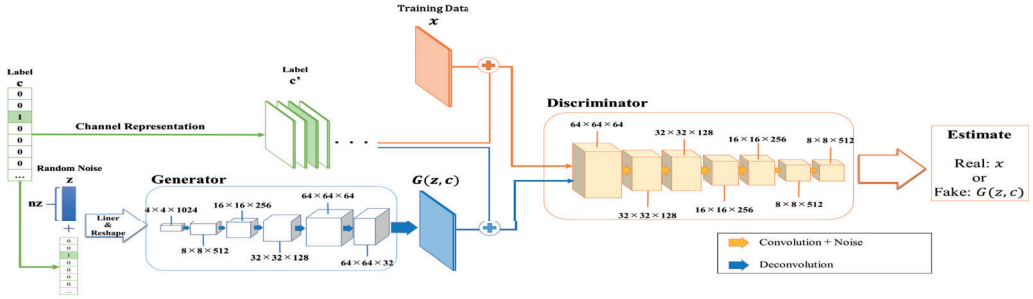


Figure 1. Proposed Conditional DCGAN which generates conditional handwritten characters

1. Introduction

Of late, more and more Deep Learning techniques which deal with generating images are been developed and as a result realistic images are being generated. In addition to being able to generate images that are realistic, the potential of deep learning is immense, such as supplementing blank areas and learning to imitate styles of famous painters to create artistic images.

To further test the potential of deep learning, we tried generating handwritten characters by developing a model called *Conditional DCGAN* which combines DCGAN with cGAN (Conditional GAN). A conditional label is added as additional input along with the image input to the model. The target kinds of character we dealt with in this study are not only digits, but also alphabets and katakana (Japanese script) and some special characters. The goal of this study is generating handwritten characters by making a distinction among more than 90 different kinds of them.

In our experiment, we obtained results by changing the dimensions of the random noise which is part of the input for the models. It can be inferred from our results that when the number of dimensions of noise falls below the number of labels, the model cannot generate images that are likely to be characters; and on the other hand, if the number of types exceeds 90, the model can generate the specified characters.

Contact: Rina Komatsu, Faculty of Science & Technology, Sophia University, Tokyo, Japan, r_komatsu@outlook.com

2. Related Work

We construct the model (the architecture shows in Figure 1) based on the techniques from DCGAN and cGAN. This section introduces generating method: GAN and DCGAN, also introduce cGAN to generate conditional images.

2.1 GAN & DCGAN

GAN: Generative Adversarial Net [Ian J. Goodfellow, 2014] is a generative network model that generates images by training a Generator and a Discriminator that are tied together in an adversarial relationship. The Generator plays the role of generating images from a given probability density distribution with random noise input, while the Discriminator plays the role of distinguishing the real input data from the fake data generated by the Generator. However, GAN has the weak points that the probability density distribution Generator learns is unable to indicate clearly and training Generator and Discriminator tend to unstable [Naoki Shimada et al, 2017].

DCGAN: Deep Convolutional Generative Adversarial Network [Alec Radford et al, 2015] is a generative model designed to solve this weak point by employing stable learning techniques such constructing fractional-strided convolution in Generator and strided convolution in Discriminator, in addition, instead of pooling layers, adapting batch normalization [Sergey

Ioffe et al, 2015] to each layer and so on. The Generator in DCGAN extends the information through upsampling from random noise input, while the Discriminator extracts feature maps through convolutions. As a result, DCGAN succeeds in generating more realistic images than GAN.

About how to calculate Loss GAN and DCGAN utilize Discriminator's output in loss function shown in formulae (1) and (2) to update each the parameters of each model. Formula (1) is loss function for Discriminator and (2) is the one for Generator. If the Discriminator learns good work in distinguishing, then log(D(x)) increases and 1-log(G(z)) decreases on the contrary. On the other hand, if the Generator reaches a matured level that deceives the Discriminator, then log(D(z)) increases.

$$L_D(G,D) = E\big[\log\big(D(x)\big)\big] + E[1 - \log(D(G(z)))] \quad (1)$$

$$L_G(G,D) = E[-\log(D(G(z)))] \quad (2)$$

where,
x is the training sample data, and
G(z) is the Generator's output from random noise z.

2.2 cGAN: Conditional GAN

cGAN: Conditional GAN [Mehdi Mirza et al, 2014] is the generative model which can output designated images by adding auxiliary information (represented as one-hot vector) such as a label corresponding with the kinds or modality to Generator after finishing training in the Generator and Discriminator.

Figure 2 shows a simplified structure of Conditional Adversarial Nets dealing with the auxiliary information. Random noise z and an auxiliary information y are input to the Generator combined forward to hidden layer. These jointed data help the Generator to suggest the probability density distribution to which the training sample data belongs. Also, training sample data *x* or generated ones G(*z*|*y*) and *y* are input to Discriminator combined in same.

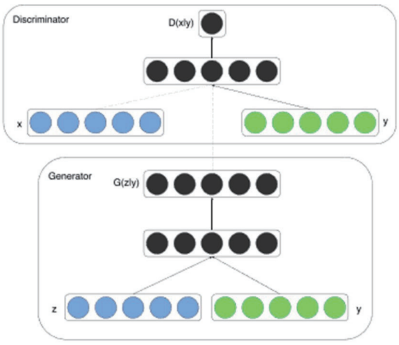


Figure 2. Simple Structure of Conditional Adversarial Net (adapted from Figure 1 in [Mehdi Mirza et al, 2014])

2.3 Conditional DCGAN (Constructed model in this study)

Figure 1 is the structure of proposed model in this study through trial and error finding stable training between Generator and Discriminator relatively quickly. In section 4 "Result" introduces the result using the Generator in this architecture.

Explaining details in this proposed model, Generator and Discriminator have the common factor that there is additional input named the auxiliary information represented one -hot (In this study, the auxiliary information is replaced to the label information related with the kind of characters). In Generator, the random noise z which consists of the number of dimensions nz and label information *c* are merged and input to Liner layers, then proceeded to up-sampling as output G(*z*,*c*) by deconvolution. In Discriminator, c is transposed to channel representation c' since Discriminator's input is represented with channel like training sample data x or G(*z*,*c*), then merged them and extract feature maps by convolution. When proceeded to convolution, input in each layer is added some noise for stable learning [Martin Arionsky et al, 2017].

Generator, Discriminator Loss is obtained by using the output from Discriminator like GAN and DCGAN. Formula (3) is loss function for Discriminator and (4) is the one for Generator.

$$L_D(G,D) = E\big[\log\big(D(x,c)\big)\big] + E[1 - \log(D(G(z,c),c))] \quad (3)$$

$$L_G(G,D) = E[-\log(D(G(z,c),c))] \quad (4)$$

3. Experiment

3.1 Handwritten character dataset

As the target for handwritten character dataset, we used ETL-1 Character Database [Electrotechnical Laboratory, 1973-1984] from Electrotechnical Laboratory (succeeding organization: National Institute of Advanced Industrial Science and Technology).

In the ETL-1 Character Database, the handwritten character images are grayscale and have a unified size of 64×63. The dataset contains 96 different characters: 10 Arabic numerals, 26 large alphabets, 12 special characters and 48 katakana letters. These handwritten characters were collected from 1445 writers, by making each writer write one character at a time on an OCR sheet. The total number of samples collected were 141,319.

In the training process of the Generator and the Discriminator, we treated this dataset as training sample data x.

3.2 Experiment Environment

The training of the Generator and Discriminator to distinguish 96 different kinds of characters is implemented in the Python programming language and Chainer deep learning library [Seiya Tokui et al, 2015]. We also used NVIDIA GeForce GTX 1080 Ti graphic boards to speed up the training as much as possible.

3.3 Experiment Setup

As an initial setting, the whole training sample images are resized to 64×64 and set the weight decay parameter λ= 0.00001.

The following steps count as 1 epoch. We repeated training the Generator and Discriminator for 100 epochs, every time employing a minibatch size 50.

Step 1:
This step consists in preparing the Generator's input, random noise and the label information. Random noise *z* is generated from uniform random distribution in the range [-1, 1], setting the number of dimensions as *nz*. Label information c is represented

with one-hot vector corresponding to the ID related to each type of character. The data shape of c becomes (batch size, label num, 1).

Step 2:

The z and c inputs are merged into the Generator to generate the output data G(z, c).

Step 3:

To the Discriminator, G(z,c) as fake data is input merged with c' which is represented in channel from c (The data shape of c' becomes (batch size, label num, h, w)). Next, the training sample data x is input merged with c'.

Step 4:

From the Discriminator's output, Generator and Discriminator Loss is calculated and the relevant parameters are updated in each model. As an optimization function, we employed the Adam function [Diederik P. Kingma et al, 2014]. The Adam function parameters in the Generator and Discriminator network models which assisted stable training in our study are shown in Table 1.

Table 1: Adam function parameters

Parameters	α	β_1
Generator	0.001	0.5
Discriminator	0.0002	0.5

4. Result

Using the Generator in our Conditional DCGAN, this section introduces the generated result changing *nz* =32,64,96(corresponding to the number of character kinds), 256,1024 and 4096(same to whole image size we set).

4.1 Generating conditional handwritten characters

To make sure the Generator output handwritten characters designating c, we prepared 5 kinds of characters. Figure 3 shows each kind of handwritten character image picked up from training sample data.

Figure 3. The targets for generating (picked up from ETL-1 Character Database)

Figure 4 is the result generated by using Generators of varying nz values. In the images depicted in Figure 4, vertical axis means the output changing label information and horizontal axis means the output using different random noise z.

From the results in *nz*=32,64 and 96, we can see that there are outputs which are likely handwritten characters, but they do not reflect the label information. Most output were the handwritten character not belong to the kind in training sample data. Also, same images are generated although changing z.

On the other hand, in Generator with nz set to 256, 1024, 4096, it is possible to generate by reflecting the designation of target character type. Thus, there is no confusion between similar characters such as "8" and "S", "シ" and "ツ" which are similar in shape. Moreover, in the result of changing the random noise, it was possible to generate an image in which its peculiarity appeared rather than a similar image, such as when the character is large or small, or the thickness of the line is different. Moreover, it was able to generate distinct characters despite the size being smaller (*nz* = 256; image size: 64 × 64).

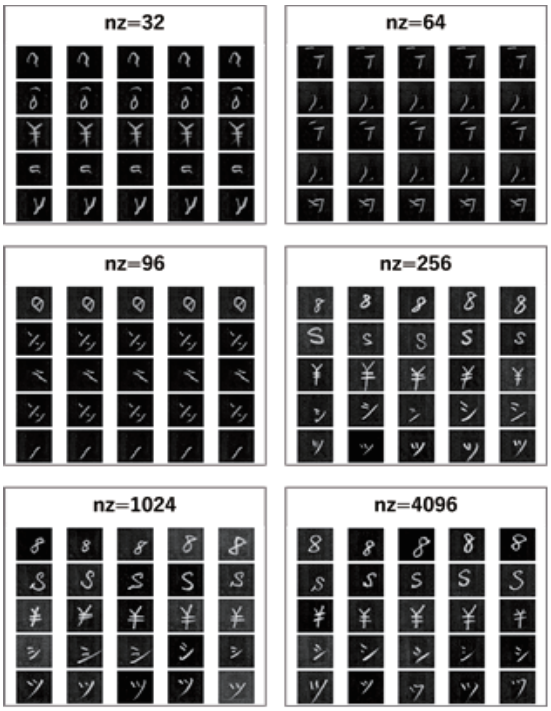


Figure 4. Conditional output from Generator changing nz

4.2 Loss changes in Generator and Discriminator

The Loss specific to the Generator and Discriminator for each epoch is shown in Figure 5.

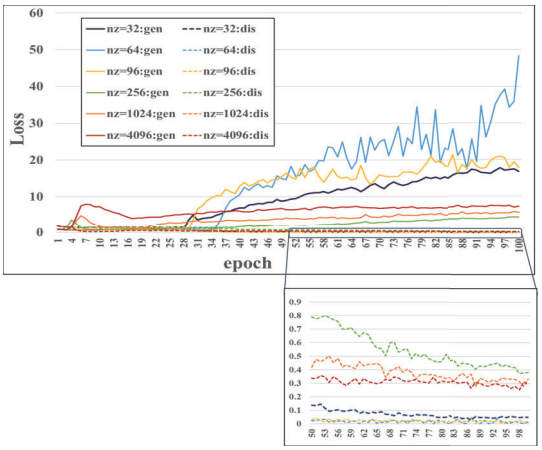


Figure 5. Loss changes from epoch 1-100.

As can be seen in Figure 5, in the model *nz* = 32,64,96, as the epochs progressed, the Generator Loss steadily increased, while the Discriminator loss gradually decreased near to 0. The difference in loss between Generator and Discriminator at epoch wider than the ones in conditional image generation. This result implies that gradient vanishing occurred in the Generator since Discriminator learned to distinguish between the real data and the fake data much before the Generator optimized to deceive the matured Discriminator [Ian Goodfellow, 2016].

Figure 6 shows the result of generating the whole of 96 kinds of characters as the target we set, from a larger *nz*=4096 to a smaller *nz*=256. Each Generator could output almost all kinds of handwritten images, making distinction just by changing the label information.

It can be inferred from our results that when the number of dimensions of noise falls below the number of labels, the model cannot generate images that are likely to be characters; on the other hand, if the number of types exceeds 90, the model can generate the specified characters.

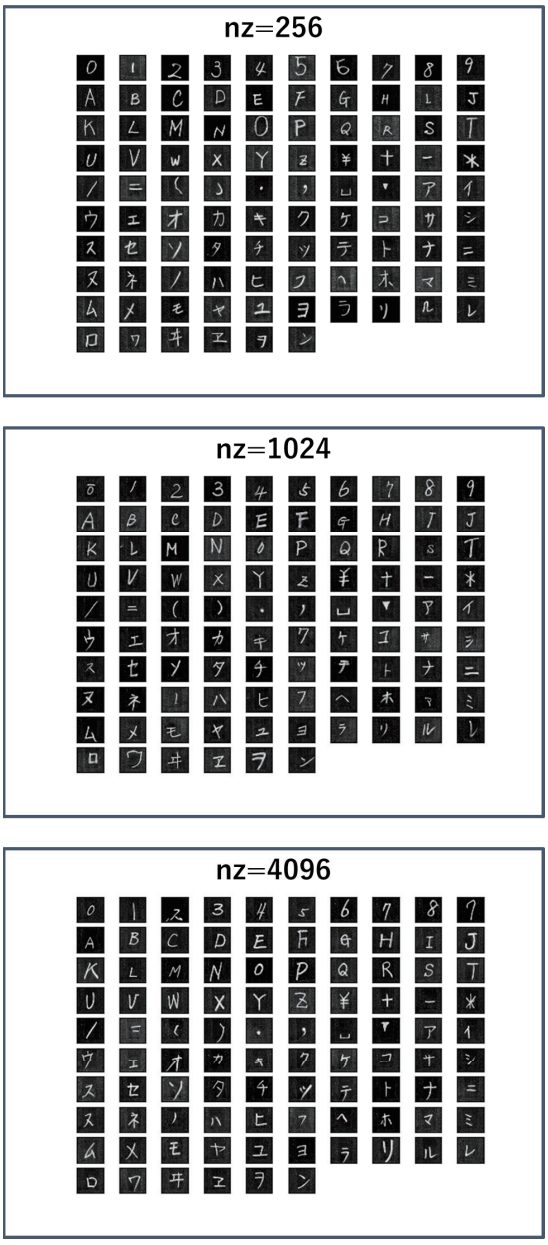


Figure 6. Handwritten characters produced by Generator

5. Conclusion and Future work

In this study, to be able to generate handwritten characters distinguishing among 96 different kinds of characters by adding UI designation, we constructed Conditional DCGAN. This model adapted DCGAN techniques using deconvolution for up-sampling at the Generator and convolution for extracting feature maps, and cGAN technique that adds label information to Generator and Discriminator. Through training our Generator and Discriminator with the dimension of random noise over the kinds, Generator could output the entire set of characters as a result.

In our future work, since the data shape of label information at the Discriminator in Figure 1 is (batch size, label num, h, w), large load will be applied to the model if dealing with over thousand kinds of characters like kanji. To solve this problem, constructing more compact Discriminator so that Discriminator's label information could keep the shape same as the one generated by the Generator and compressed through linear function. We want to try generating conditional images making distinction among over thousand kinds of images with compact Conditional DCGAN as the next challenge.

References

[Ian J. Goodfellow et al, 2014] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville & Yoshua Bengio: Generative Adversarial Nets, Advances in neural information processing systems, pp. 2672-2680, 2014.

[Naoki Shimada et al, 2017] Naoki Shimada & Takeshi Ooura: INTRODUCTION TO DEEP LEARNING WITH Chainer, Gijutsu-Hyohron Co (Japan), 2017.

[Alec Radford et al, 2015] Alec Radford, Luke Metz & Soumith Chintala: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, arXiv preprint arXiv:1511.06434, 2015.

[Sergey Ioffe et al, 2015] Sergey Ioffe & Christian Szegedy: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, arXiv preprint arXiv:1502.03167, 2015.

[Mehdi Mirza et al, 2014] Mehdi Mirza & Simon Osindero: Conditional Generative Adversarial Nets, arXiv preprint arXiv:1411.1784, 2014.

[Martin Arjovsky et al, 2017] Martin Arjovsky & Léon Bottou: Towards Principled Methods for Training Generative Adversarial Networks, arXiv preprint arXiv:1701.04862, 2017.

[Electrotechnical Laboratory, 1973-1984] Electrotechnical Laboratory: Japanese Technical Committee for Optical Character Recognition, ETL Character Database, 1973-1984.

[Seiya Tokui et al, 2015] Seiya Tokui, Kenta Oono, Shohei Hido & Justin Clayton: Chainer: a Next-Generation Open Source Framework for Deep Learning, Proceedings of workshop on machine learning systems (LearningSys) in the twenty-ninth annual conference on neural information processing systems (NIPS). Vol. 5, pp. 1-6,2015.

[Diederik P. Kingma, 2014] Diederik P. Kingma & Jimmy Lei Ba: Adam: A Method for Stochastic Optimization, arXiv preprint arXiv:1412.6980, 2014.

[Ian Goodfellow, 2016] Ian Goodfellow: NIPS 2016 Tutorial: Generative Adversarial Networks, arXiv preprint arXiv:1701.00160, 2016.

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Sparse Damage Per-pixel Prognosis Indices via Semantic Segmentation

Takato Yasuno[*1]

[*1] Research Institute for Infrastructure Paradigm Shift (RIIPS)

Efficient inspection and accurate prognosis are required for civil infrastructures with more than 30 years since completion. If we can detect damaged photos automatically per-pixels from the record of the inspection record and countermeasure classification of drone inspection vision, then it is possible that countermeasure information can be provided more flexibly, whether we need to repair and how large the expose of damage interest. A piece of damage photo is often sparse as long as it is not zoomed around damage, exactly the range where the detection target is photographed, is at most only one percent. In this paper, we propose three damage detection methods of transfer learning which enables semantic segmentation in an image with low pixels using damaged photos of drone inspection. Furthermore, we propose prognosis indices to make a decision repair-priority such as the counts index of pop-outs region and the per-pixel area counts index of each pop-out based on morphology image processing. In fact, we show the results applied this method using the 40 drone inspection images whose size is 6,000 x 4,000 on an infrastructure, where each image is partitioned into 400 crops, so the total number of input images is 16,000 for training deep neural network. Finally, future tasks of damage detection modeling are mentioned (211words).

1. Introduction

Deterioration of civil engineering structures is progressing in recent years, including a large number of concrete structures. Improving efficiency of scheduled inspections is a pressing issue, since the cost of inspections comprises a large proportion of maintenance costs for local governments, which are also experiencing manpower shortage for technical personnel. There are often opportunities to apply deep learning as a method for improving efficiency of inspections on social infrastructure and studies have been conducted on this issue. Dam general inspection is required for dam once every 30 years and as a result, images of damage have been accumulating (Ministry of Land, Infrastructure, 2013). If it were possible to utilize images of damage that are attached to inspection reports, data from scheduled inspections from past years can be input for the purpose of deterioration learning. If it could be possible to automatically calculate numerical scores for the extent of damage based on images of damage, this would be useful in deciding whether any repairs work should be performed and for setting the order of priority among candidates for repairs. There are past studies on detecting cracks in concrete on bridges, structures, plants, etc.

Especially, for dam structural health monitoring, it is important to prognosis pop-outs owing to be greater impact on the health of dam embankment. In area of low quality aggregates, as a result of the water absorption of the concrete, the soft stone having a high water absorption rate becomes saturated. When the freezing temperature is reached, pressure due to volume expansion occurs. However, the detection model for pop-out is only at its incipient stages, so it would be difficult to claim that this is an established means for concrete damage deterioration learning and prediction. This paper proposes a practical method applies semantic segmentation (segmentation) of concrete damage using images of damage from drone-base inspections. Results are shown from actually applying this method on sparse images of damage, focusing on images of pop-outs among images of damage to dam embankment. Finally, references will be made to issues of damage detection modeling as well.

Contact: Takato Yasuno, RIIPS on 5-20-8, Asakusabashi, Taito-ku, Tokyo, 111-8648, tk-yasuno@yachiyo-eng.co.jp

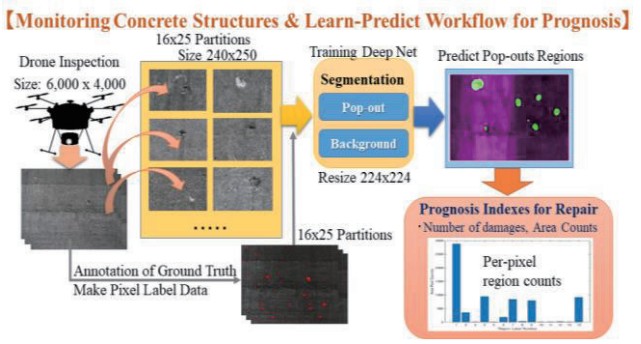


Figure 1: Monitoring concrete structures from drone-base inspection to train segmentation networks and damage prediction for prognosis indices.

2. Related Studies and Damage Images

2.1 Damage detection studies for civil infrastructures

Since 2002, there has been an accumulation of studies (Wu, 2002) (Chun, 2015) on resolving damage detection using neural networks (ANN) for the purpose of continuous surveillance of bridges. Many instances of damage detection modeling for machine learning have been conducted over the past 15 years, including the ANN, as well as the PCA, SVM, GA and other such solution methods (Gordan, 2017). Since the potential of convolutional neural networks (CNN) to exhibit high degrees of accuracy in classifying one million images into 1,000 classes was reported in 2012 (AlexNet, 2012), there has been active reporting of studies on solution methods of the CNN, which provides solutions with greater accuracy than conventional methods for label categorization of overall images, object detection and semantic segmentation at the pixel level. There have been a number of studies conducted on damage classification of at the whole-image level for cracks and corrosion of road pavement, structures and bridges, for detection of damage to civil engineering structures (Gopalakrishnan, 2018) (Ricard, 2018), as

well as damage segmentation at the pixel level (Hoskere, 2017). A report was made on a study that applied deep CNN to conduct four classes of damage segmentation, namely no damage, only separation, exposure of rebar (with and without rust), using 734 images of damage (Guillamon, 2018). The breakdown of the damage classes, however, indicated a distribution biased to the third class, for which there were 510 images, and as such, distortion in the training images cannot be denied. Dimensions of the images of damage were widely varied, being 640 x 480, 1,024 x 768, and 1,600 x 1,200. The potential for learning with the index that represents the degree of matching between prediction and reality, mIoU (class mean IoU) to the level of 0.6 to 0.8 was indicated by using some types of CNN models for fully convolutional networks (FCN) in entering images of such diverse dimensions. The use of the damage detection modeling that utilizes solution method of CNN, however, has just been started and as such, it would be difficult to claim that this is an established general-purpose method for damage detection in management of infrastructure. This paper proposes a practical method for damage segmentation with considerations for sparse characteristics of damage images from drone-base inspections. Furthermore, using the output of prediction RGB-images by the trained semantic segmentation, we propose two morphological indices such as the number of identified damages and the per-pixel counts of each damage region for prognosis to make a decision repair-priority.

Table 1: Comparison of the per-pixel counts between the target pop-outs region and the background region.

Example consisting of 40 damage drone inspection images	Total number of pixels per damage image	The number of pixels per image	Percentage per image
Background	954,339,801	23,858,495	99.4%
Damage to region of interest (ROI)	5,660,199	141,505	**0.6%**
Total per image	960,000,000	24,000,000	100.0%

2.2 Characteristics of Damage Images

This paper provides a practical observation on characteristics of images of damage, using 40 images of damage in which pop-out has been captured through drone-base inspection of dam embankment, whose size is 6,000 x 4,000. While generality cannot be guaranteed with these characteristics, they are considered to lead the way for utilizing images of damage. Characteristics of general conditions and damage for pop-out is as follows (CERI, 2016). Pop-out is a crater-like indentation generated by destruction due to the expansion of aggregate particles on the concrete surface. These are often observed in aggregates with high water absorption and in poor quality. Pop out is the meaning of “jumps out suddenly”. In the case of low-quality aggregate, as a result of the water absorption of the concrete, the soft stone having a high water absorption rate becomes saturated. At this time, when the freezing temperature is reached, pressure due to volume expansion occurs, the surface portion peels off, and then a crater-like hole is formed.

Table 1 shows the summary value for the damage area (region of interest: ROI) subject to detection, as well as other regions in the background, counted at pixel level. No advance manipulation was conducted on images to unify photographing distance and picture quality. The number of pixels per image was 24 million pixels. The proportion of these that include targeted damage was only 0.6%. The first characteristic of damage image is the sparsity of the area comprised of ROI.

3. Per-pixel Learning and Prognosis Indices

3.1 Damage segmentation for prediction

The FCN-Alex (Long, 2015), as well as the SegNet-VGG16 (Badrinarayanan, 2016) are compared where appropriate, as a method for learning transfers of semantic segmentation. The solution method used in this paper by itself does not present any innovation but the extremely sparse proportion of detection target ROI on any given image is a characteristic and the intention was to derive a practical method that can be applied to images of damage with sparse pixel labels. The FCN-Alex is a transfer learning of AlexNet and the CNN is implemented to the deepest layer, making it a deep neural net (DNN) of 23 layers in depth. Learning is possible with relatively short calculation time and prediction output for exhaustive detection of targeted damage can be achieved. SegNet-VGG16 is a method of transfer learning used to identify objects for automatic driving and a DNN with depth of 91 layers.

This paper applies the four deep neural networks described above to images of damage to compare calculation execution time, accuracy and prediction output image. There is a problem of no improvements being evident with loss functions when the SGDM is used in the optimization method for hyper parameters, as gradients of the detection target are eliminated due to the sparse characteristic of the damage image. In order to overcome this issue, the gradient of the detection target is captured with good sensitivity and the previously updated quantities are deleted where appropriate, and the RMSProp, which has a characteristic formula for error function that eliminates the amount of change in gradients of detection targets by taking square root of the amount of change in gradient, is adopted (Hinton, 2012) (Mukkamala, 2017). The weighting factor for the updating amount was set to 0.99. The learning coefficient for the overall model was set to 1E-5 and the minibatch was set to 32.

3.2 Morphological indices for prognosis

The word morphology commonly denotes a branch of biology that deals with the form and structure of animals and plants. We use the same word here in the context of mathematical morphology as a toll for extracting image components that are useful in the representation of region shape. We are interested also in morphological techniques for pre- or post-processing, such as morphological filtering, thinning, and pruning. In image-processing applications, dilation and erosion are used most often in various combinations (Serra 1992; Gonzalez 2008). This paper proposes some prognosis indices to make a decision repair-priority such as the counts index of pop-outs region and the per-pixel area counts index of each pop-out based on morphological image processing, such as dilation and erosion operation.

On the prediction of pop-out damage segmentation, there are some extremely small size of pop-outs, so that we may overlook

them. Also, the shape of pop-outs are not always like circle, but these are complex shape or partially connected with various size of pop-outs. This paper proposes two practical morphological operation, dilation and erosion, in terms of the union (or intersection) of an image with a translated shape called structuring element. At first, we translate the prediction RGB-image of pop-out segmentation into a binary mask image with pop-out foreground (1-valued pixel, white color) and with background (0-valued pixel, black color). Dilation is an operation that "grows" or "thickens" objects in the extremely small images of pop-outs. This growing is controlled by a shape referred to as a structuring element, such as linear, disk, octagon etc. This paper proposes the disk-shaped structuring element with radius r=3. Further, erosion "shrinks" or "thins" objects in a binary image like complex shape and partially connected with various size of pop-outs. This paper proposes these morphological operations applied to the masked prediction of pop-outs images. By these operation, it is possible to avoid overlooking the small pop-outs, and we can extract the complex shape or partially connected pop-outs, in order to count the number of pop-out and the each region pixel size more accurately and efficiently.

4. Applied Results

4.1 Training results

The input data was 40 images whose size is 6,000 x 4,000 from drone-base inspections of dam embankment. In order to bring them closer with the input size of deep pre-trained network, we partitioned each original image into 25 x 16 equal 400 crops whose size was 250 x 240. The usage rate of the training and test data was set to Train: Test = 99:1. The transition of loss function in the learning process applied to the pop-out segmentation is shown in Figure 2. The calculation conditions are 490 cycles per epoch for a total of 24,500 repeated calculations in 50 epochs. The loss value of the FCN-AlexNet is transitioning at a lower level than SegNet-VGG16. This FCN models, however, have large dispersion of loss values and their disadvantage is that they make for unstable learning processes. The loss function of the SegNet-VGG16 does not offer minimum values, but up and down fluctuations remain small early on, which can be interpreted to offer superior stability for the learning process.

Table 2: Comparison of indices for pop-out segmentation models.

DNN model	Time calculation	Mean mIoU	Weighted wIoU
FCN-AlexNet	466min.	0.5811	0.9861
SegNet-VGG16	832min.	**0.5967**	0.9856

Table 2 shows the calculation time, accuracy, mean-IoU and weighted-IoU index of respective segmentation model. The FCN-AlexNet offers a relatively short calculation time of 466 minutes. This net achieved the index such as mIoU = 0.5811, and wIoU = 0.9861. Meanwhile, the SegNet-VGG16 offers about two times calculation time compared with FCN-AlexNet, and indicates the score of mIoU of 0.5967 and wIoU of 0.9856. While each weighted wIoUs have almost no difference, but regarding the score of the mean mIoU the SegNet-VGG16 is superior with the FCN-AlexNet to select better pop-out predictor.

4.2 Prediction results

Figure 2 shows an output RGB-image of predictions for a test image whose size is 600 x 400, using the trained SegNet-base predictor of pop-out segmentation. Here, the region of prediction are shown in green color. In contrast, the region of background are shown in magenta color. Figure 3 shows the translated binary mask with pop-out foreground (1-valued) and with background (0-valued). We operated the morphological operations applied to the masked prediction of pop-outs images. Further, we compute the centroid of each pop-out region and set the pop-out number in order to represent the counts index accurately, here the total count of pop-outs is 14. Figure 4 shows the per-pixel counts of each region based on the morphological image pre-processing. Figure 5 shows the bar chart that we can visualize the volume indices of pop-outs and it is possible to compare the largest size, middle size, and extremely small size of pop-outs in order to make a decision of repair-priority for infrastructure manager.

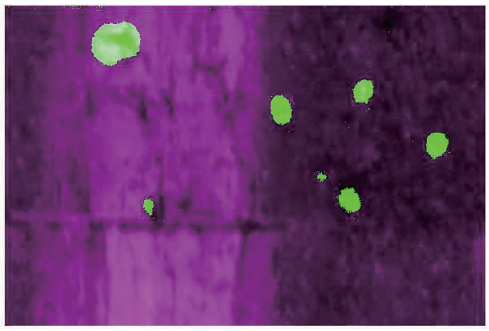


Figure 2: Trained SegNet-base prediction of pop-outs (RGB image) Here, green indicates prediction, magenta denotes background.

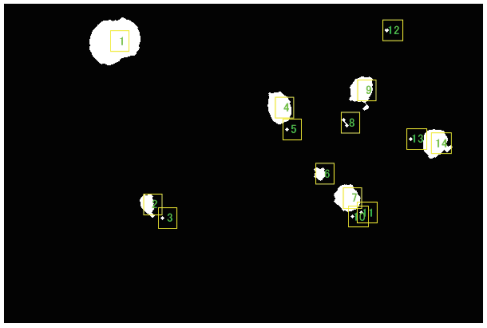


Figure 3: Counts index of identified pop-out centroid based on morphological operations with dilation and erosion.

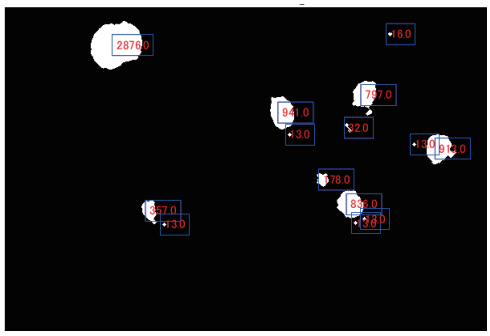


Figure 4: Per-pixel counts of each pop-out prediction region based on morphological image processing.

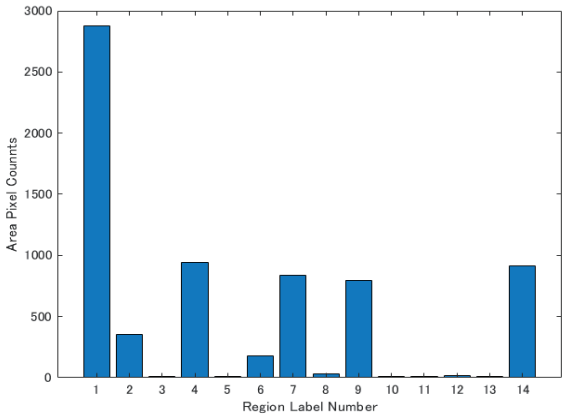


Figure 5: Area pixel count index of each pop-out for prognosis measure indices to make a decision regarding repair-priority.

5. Conclusion

5.1 Concluding remarks

This paper proposed a method for detecting pop-out by semantic segmentation, using images of damage obtained from drone-base inspections. In fact, we show the results applied this method using the 40 drone inspection images at a dam embankment, where each image is partitioned into 400 crops, so the total number of input images is 16,000 for training deep neural network. Based on transfer learning, per-pixel higher accurate prediction is possible, even to sparse damage images whose pop-out ratio per-pixel is only one percent compared with the background. The SegNet-VGG16 exhibited the better accuracy and achieved class mean mIoU index of 59.67% and weighted index wIoU of 98.56%. Furthermore, we demonstrated to compute some morphological indices, such as the counts index of identified pop-outs centroid and the per-pixel area counts index of each pop-out region for prognosis to make a decision repair-priority more accurately and efficiently.

5.2 Future works

The scope of this paper was the segmentation of pop-out for prognosis, using images from drone-base inspection of dam embankment. Monitoring various damages for standard dam inspection prescribes concrete crack, scaling, pop-outs, water leak, efflorescence etc (Ministry of Land, Infrastructure, 2013). In contrast, creation of dataset for training and prediction of various segmentation models for being predictive diagnosis before occurred pop-outs, such as “crack” and “scaling” are the issue for health monitoring. Infrastructure manager administrates a lot of aging structures other than dam well. Learning of damage segmentation models using a diverse range of images for a wide variety of other infrastructures will be the issue for the future. Predictor of damage segmentation intelligence created from scratch, i.e. U-Net by data mining accumulated images is also a challenging issue. Furthermore, 3-dimension segmentation is more useful for volume counting to measure the depth of damage.

[Acknowledgments] We would like to express our gratitude to obtain practical information about Deep Learning and Image Processing Toolbox for training and prediction from Mr. Shinichi Kuramoto and Mr. Takuji Fukumoto.

References

[Ministry of Land, Infrastructure 2013] Water Management Land Conservation Bureau River Environment Division : Dam General Inspection Procedure Commentary, 2013.

[CERI 2016] Public Works Research Institute CERI : Survey and Countermeasure Guidance on the Structure Suspected of Frost Damage (draft), 2016.

[Wu 2002] Wu, Z., Xu, B. Yokoyama, K. : Decentralized Parametric Damage Detection Based on Neural Networks, Comput. Civ. Infrastruct. Eng., 17, pp.175-184, 2002.

[Chun 2015] Chun, P., Yamashita, et al. : Bridge Damage Severity Quantification using Multipoint Acceleration Measurement and Artificial Neural Networks, 2015.

[Gordan 2017] Gordan, M., Razak, H.A. et al. : Recent Development in Damage Identification of Structures using Data Mining, Latin American Journal of Solids and Structures, pp.2373-2401.

[AlexNet 2012] Krizhevsky, A. et al. : ImageNet Classification with Deep Convolutional Neural Networks, NIPS, 2012.

[Gopalakrishnan 2018] Gopalakrishnan, K., Gholami, H. et al. : Crack Damage Detection in Unmanned Aerial Vehicle Images of Civil Infrastructure using Pre-trained Deep Learning Model, International Journal for Traffic and Transport Engineering, 8(1), pp.1-14, 2018.

[Ricard 2018] Ricard, W., Silva, L. et al. : Conclete Cracks Detection based on Deep Learning Image Classification, MDPI Proceedings, 2, 489, pp.1-6, 2018.

[Hoskere 2017] Hoskere, V., Narazaki, Y. et al : Vision-based Structural Inspection using Multiscale Deep Convolutional Neural Networks, 3rd Huixan International Forum on Earthquke Engineering for Young Researchers, 2017.

[Guillanmon 2018] Guillamon, J.R. : Bridge Structural Damage Segmentation using Fully Convolutional Networks, Universitat Politecnica de Catalunya, 2018.

[Long 2015] J. Long, E. Shelhamer, T. Darrell : Fully Convolutional Networks for Semantic Segmentation, CVPR, pp3431-3440, 2015.

[Badrinarayanan 2016] V. Badrinarayanan, A. Kendall, et al., SegNet: Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, ArXiv:1511.00561v3, 2016.

[Hinton 2012] G. Hinton, N. Srivastava, K. Swersky : Lecture 6d – A Separate, Additive Learning Rate for Each Connection, Slides Lecture Neural Networks for Machine Learning, 2012.

[Mukkamala 2017] M.C. Mukkalmala et al.: Variants of RMSProp Adagrad with Logarithmic Regret Bounds, 2017.

[Serra 1992] Serra,J., Vincent,L.: An Overview of Morphological Filtering, Circuits, Systems and Signal Processing, 11(1), pp47-108, 1992.

[Gonzalez 2008] Gonzalez, R.C., Woods, R.E. : Digital Image Processing, 3rd ed., Prentice Hall, 2008.

[Yasuno 2018] T. Yasuno: Infra Machine Learning for Predictive Maintenance via Classification Models, 32th Journal of Society for Artificial Intelligence, 3Z1-04, 2018.

[Yasuno 2019] T. Yasuno, M. Amakata, J. Fujii, Y. Shimamoto : Color-base Damage Feature Enhanced Support Vector Classifier for Monitoring Quake Image, IAPR: CCIW, 2019.

(2019.Feb.15)

International Session | International Session | [ES] E-2 Machine learning

[3B4-E-2] Machine learning: social links

Chair: Lieu-Hen Chen (National Chi Nan University), Reviewer: Yasufumi Takama (Tokyo Metropolitan University)

Thu. Jun 6, 2019 3:50 PM - 5:30 PM Room B (2F Main hall B)

[3B4-E-2-01] Social Influence Prediction by a Community-based Convolutional Neural Network

Shao-Hsuan Tai[1], Hao-Shang Ma[1], ○Jen-Wei Huang[1] （1. National Cheng Kung University）

3:50 PM - 4:10 PM

[3B4-E-2-02] A Community Sensing Approach for User Identity Linkage

○Zexuan Wang[1], Teruaki Hayashi[1], Yukio Ohsawa[1] （1. Department of Systems Innovation, School of Engineering, The University of Tokyo）

4:10 PM - 4:30 PM

[3B4-E-2-03] Learning Sequential Behavior for Next-Item Prediction

○Na Lu[1], Yukio Ohsawa[1], Teruaki Hayashi[1] （1. The University of Tokyo）

4:30 PM - 4:50 PM

[3B4-E-2-04] Application of Unsupervised NMT Technique to Japanese--Chinese Machine Translation

○Yuting Zhao[1], Longtu Zhang[1], Mamoru Komachi[1] （1. Tokyo Metropolitan University）

4:50 PM - 5:10 PM

[3B4-E-2-05] Synthetic and Distribution Method of Japanese Synthesized Population for Real-Scale Social Simulations

○Tadahiko Murata[1], Takuya Harada[1] （1. Kansai University）

5:10 PM - 5:30 PM

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Social Influence Prediction by a Community-based Convolutional Neural Network

Shao Hsuan Tai[*1] Hao-Shang Ma[*2] Jen-Wei Huang[*3]

[*1*2*3]Institute of Computer and Communication Engineer,
Department of Electrical Engineering,
National Cheng Kung University, Tainan, Taiwan

Learning social influence between users on social networks has been extensively studied in a decade. Many models were proposed to model the microscopic diffusion process or to directly predict the final diffusion results. However, most of them need expensive Monte Carlo simulations to estimate diffusion results and some of them just predict the size of the spread via regression techniques, where people who will adopt the information becomes unknown. In this work, we regard the prediction of final influence diffusion results in a social network as a classification problem to avoid expensive simulations with knowing the final adopters. We first address the problem on a deep neural network and utilize the diffusion traces to train the network. Furthermore, we propose a community-based convolutional neural network to capture the information of local structure with the aforementioned network. The proposed model is referred to as the Social Influence Learning on Community-based Convolutional Neural Network, SIL-CCNN. In the experiment, SIL-CCNN shows the promising results in both synthetic and real-world datasets.

1. Introduction

Nowadays people tend to share their life, emotions, and opinions to others on social network websites. Two representative diffusion models, Independent Cascade, IC, model and Linear Threshold, LT, model, were reformulated by Kempe [3]. However, there are several limitations of predicting the information diffusion using diffusion models. The influence probabilities between users and the active threshold of a user should be measured or learned from many personal features such as users' preferences and different relationships. In real social networks, the features are not easy to extract since the data sometimes is not complete. In addition, to get the results of IC/LT model, we need to conduct a huge number of simulations.

Actually, the prediction of the information diffusion process and the final diffusion results models can be regarded as a classification problem. Given the information sources and the network structure, the individuals are classified into active or inactive classes in the final diffusion result. The active class is corresponding to the individuals being influenced successfully in the information diffusion process. Some related works aim to predict the size of information spreads as a classification or regression problem [1, 7]. Different with diffusion models, these methods usually do not learn the individuals who are actually influenced by the information. We would like to know exactly who are influenced and who are not.

To overcome above limitations and solve the classification problem, we propose an influence prediction model based on community-based convolutional neural network. First, we consider the influence propagation process in the past to learn the influence between individuals. Then, most of the diffusion models consider the network structure, i.e., the relations between individuals, as their features. However, the information of the whole network structure may not be useful for the classification problem whereas the local network information of a single individual should be helpful. We try to embed the local structure of an individual into our model to learn the local relations. The community structure in social networks represents a cluster of individuals sharing connections that are stronger than those with individuals outside the community. The information diffusion in a community should be faster than outside the community. Therefor, we use the convolutional neural networks to mine the local relations within the community structure. The idea is that convolutional neural networks are good to extract the effects of a small group of individuals around an individual by extracting valuable relations from the structure. Finally, the influence traces and the community structure information are combined into our model as training features of the deep neural network. The proposed scheme is referred to as the Social Influence Learning on Community-based Convolutional Neural Network, SIL-CCNN.

The remainder of the paper is organized as follows. Works related to this work are outlined in Section 2. Section 3 details the proposed methodology. Experiment results and conclusions are presented in Section 4 and Section 5.

Contact: Jen-Wei Huang, Institute of Computer and Communication Engineer,
Department of Electrical Engineering,
National Cheng Kung University, Tainan, Taiwan.
Email: jwhuang@mail.ncku.edu.tw
Tel:(+886)-6-2757575#62347

2. Previous Work

In this section, we will briefly introduce other related works on diffusion models and the prediction of information spread size.

2.1 Diffusion Models

Diffusion models can be used to identify social influence by tracking information propagating through a social network. Nowadays, some diffusion models adopt the learning strategy to predict the activation since the technique of learning methodology has matured in these few years. Saito [6] first proposed a learning method for IC model. Wang [8] proposed a feature-enhanced approach, which considers not only temporal data in cascades but also additional features. Chou [2] proposed Multiple Factor-Aware Diffusion model, MFAD, that can consider many kinds of factors together. MFAD model adopts positive and unlabeled learning to train the classifiers for each individual.

2.2 Prediction of Information Spread Size

Different from diffusion models, the cascade prediction only focuses on predicting whether the information will become popular and widely spread. The problem is usually formulated as a classification or a regression problem to understand the size of potential influence of information.

Among the classification solutions of predicting information cascade, Cheng [1] proposed to use temporal and structural features for predicting the relative growth of a cascade size. As for the regression solutions, Tsur [7] proposed a content-based prediction model to include locations, orthography, number of words, lexicality, ease of cognitive process and emotional effect on various cognitive dimensions.

Our work aims to learn the hidden influence propagation from the data instances and the network structure directly without the huge number of diffusion simulations.

3. Methodology

In this section, we first propose an ordinary deep neural network model, SIL-DNN, for identifying traces of social influence. Second, we adopt a convolutional neural network to extract the local structure information of a network from the communities. Then, we propose Social Influence Learning on Community-based Convolutional Neural Network, abbreviated as SIL-CCNN, which combines the SIL-DNN and a community-based convolutional neural network to predict the final influential results.

Given a social network $\mathbb{G} = (\mathbb{V}, \mathbb{E})$, the node set $\mathbb{V}$ corresponds to the individuals, and $\mathbb{E}$ is the edge set indicating the relationships between individuals. Each influence trace (u, v, x) indicates that the nodes v adopt the information x and is influenced by source nodes u, where u and v are sets of nodes in $\mathbb{V}$. The architecture of SIL-DNN is presented in Fig. 1. In SIL-DNN, the number of neurons in the input layer is the same as the number of individuals in the social network $|V|$. The same setting is used in the output layer. For every trace, we put the vector of information sources as the input data of SIL-DNN and the output are the vector of influenced nodes. The input vector and output vector could be set as follows,

$$\begin{cases} y_i = 1, \text{if } i \in u \text{ for } (u, v, x) \\ y_i = 0, \qquad \text{otherwise,} \end{cases} \tag{1}$$

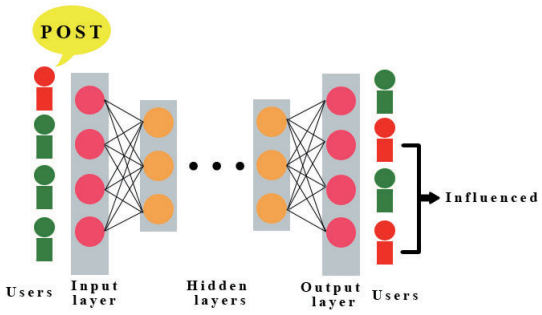


Figure 1: Social Influence Learning on Deep Neural Network Architecture

where y_i represents an individual i at the input layer. $i \in u$ for (u, v, x) indicates all traces that transmit the information x from user i. For example, an individual i provides an information x during the observation. The neuron of input layer $y_i^0 = 1$ for the information x. On the other hand, in the output layer, the $y_j = 1$ represents that node j is influenced by i and is classified as the active class. Otherwise, node j is classified into the inactive class.

3.1 Social Influence Learning on Community-based Convolutional Neural Network

In order to join the community structure to help us to predict the information diffusion results, we propose Social Influence Learning on Community-based Convolutional Neural Network, SIL-CCNN. The architecture of SIL-CCNN is presented in Fig. 2. First, we need to extract the community information in the network and form the relation matrix of each community. A list of relation matrix RM for communities in $COMM$ can be formulated. Then, we design a community-based convolutional neural network to deal with community-related information by extracting features through a convolutional layer and a pooling layer. For the input of the convolution layer, we extract the relation matrices of communities to represent the local network information of the individual. The relation matrix RM of a community is defined as follows,

$$RM = [rm_{ij}]_{d \times d}, rm_{ij} = \begin{cases} w_{ij}, (v_i, v_j) \epsilon E \\ 0, otherwise, \end{cases} \tag{2}$$

where d is the number of nodes in the community. The element rm_{ij} in the relation matrix represents the weight on the edge between node v_i and node v_j in a community. In addition, in order to account for differences in the size of communities, the community information matrices are normalized to the size of the largest community and have zero-padding.

However, if we randomly assign the order of nodes in relation matrices and put the matrices into the convolutional neural network, the small extracted region in the convolution layer would be meaningless. Therefore, we design an arrangement strategy to determine the relations of nearby

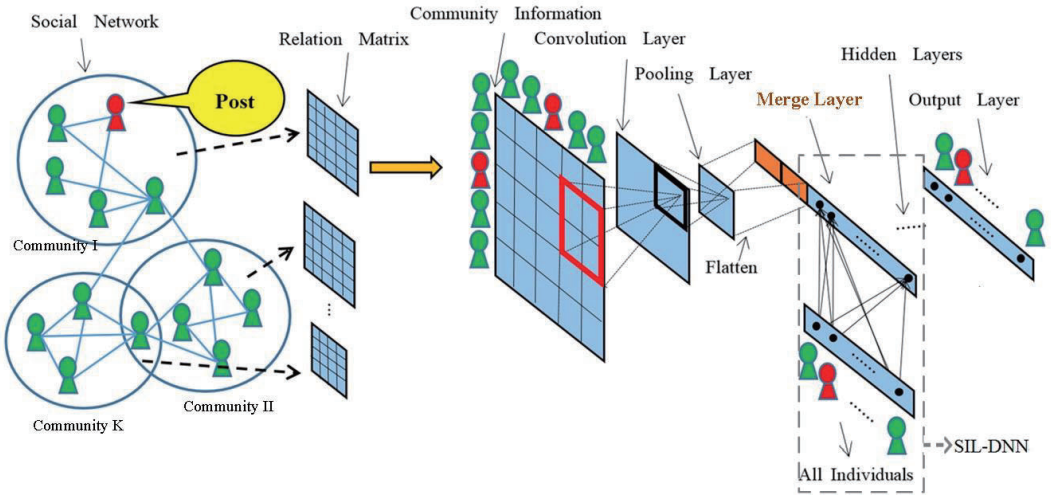


Figure 2: Social Influence Learning on Community-based Convolutional Neural Network architecture

elements. The order of individuals in each relation matrix is determined by the ranking of the number of degrees. The individual having the largest degree will be arranged to the leftmost of x-axis and the topmost of y-axis. Therefore, each block in different relation matrices shares the same weight matrix in the convolutional neural network.

In the pooling layer, we use max-pooling as our pooling function. Max-pooling is particularly well suited to the separation of features that are sparse. After the pooling layer, SIL-CCNN constructs a merge layer to combine the output of the pooling layer and the input vector of SIL-DNN. Then, several hidden layers are trained in SIL-CCNN before the output layer. Finally, a few hidden layers and one full-connected output layer are connected after the merge layer.

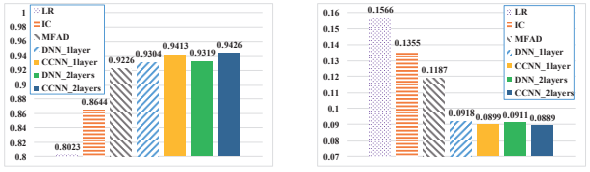
As for the detailed setting of neural network, the kernel size of the convolutional layer is defined to 3×3. In addition, we define a sigmoid function in the output layer and a cross-entropy objective function. In the training step, our goal is to minimize the following equation:

$$Q\left(W^t, W^r\right) = -\sum_{i}^{I} t_i \log y_i, \tag{3}$$

where I is the number of inputs, W^t is the weight matrices of traces, and W^r represents the weight matrices of the structure relation in SIL-CCNN. The weight matrix contains the relations between the individuals and the local structure information. The equation above is shown for one individual i at output layer k ranging over all target labels. Our objective function aims to minimize the cross entropy between the target t and the prediction y. As for the optimization, we use the well-known backpropagation algorithm and Adam [4] to compute the parameters in SIL-CCNN.

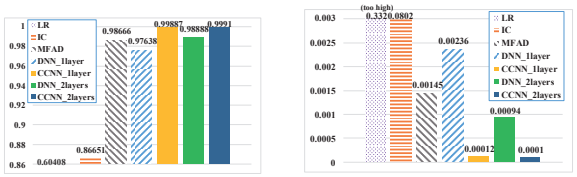
4. Experiments

In this section, we introduce experiments aiming at evaluating predicting performance in social networks using a synthetic dataset and a real-world dataset.



(a) Accuracy (b) MAE

Figure 3: Accuracy and MAE on the synthetic dataset of 5000 nodes



(a) Accuracy (b) MAE

Figure 4: Accuracy and MAE on twitter dataset

4.1 Dataset descriptions

Synthetic Data. Using the Lancichinetti-Fortunato-Radicchi (LFR) benchmark [5], we generate the synthetic dataset with 5000 nodes (2,666,674 traces). For the generation of diffusion data, we set the transmission probability uniformly between 0.1 and 1. We then choose a node at random to function as a source node and set it to be active.

Real Data. We crawl data from Twitter*1 for the period between September 2011 and May 2015. We use the experts in Healthcare Pundits*2 and Security*3 as the ini-

*1 https://twitter.com
*2 http://nursepractitionerdegree.org/top-50-health-care-pundits-worth-following-on-twitter.html
*3 http://www.marblesecurity.com/2013/11/20/100-security-

tial nodes, and crawl the followers of these experts to form a social network. The twitter data contains 20,453 users and 3,782,305 tweets. In this study, tweets were used as items, and the actions of sharing, replying, and liking are indications of the influence of a tweet.

4.2 Compared Methods

We include the following methods to predict the information diffusion results.

- **Logistic Regression** (LR). The in-degree and out-degree are included in a classifier for individual v.
- **Independent Cascade model** (IC). IC is a conventional diffusion model in which the probability of transmission from individual u to v is the ratio of items individual v adopted items from individual u.
- **Multiple Factor-Aware Diffusion model** (MFAD). MFAD is a diffusion model that can learn the social influence by multiple features [2].
- **SIL-DNN and SIL-CCNN**. We evaluated the performance of SIL-DNN and SIL-CCNN using one hidden layer (DNN_1layer, $CCNN\_1layer$) and two hidden layers ($DNN\_2layers$, $CCNN\_2layers$) in the following experiments.

4.3 Evaluated Metrics

We evaluate the performance of algorithms using the following two metrics.

- **Accuracy**. Accuracy is defined as the ratio of correct predicted answers over all answers in order to estimate the correctness of predictions.
- **Mean Absolute Error** (MAE). We computed the mean absolute error as follows: $MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n}$, where y_i is the truly adopted result of individual i, x_i is the estimated adoption probability of individual i, and n is the number of individuals in the network.

4.4 Results and Discussions

The comparison results on synthetic dataset is shown in Fig. 3. SIL-DNN and SIL-CCNN outperform the other three methods in the synthetic dataset. In addition, SIL-CCNN have higher accuracy and lower MAE than SIL-DNN using the same number of hidden layers. The results show that the information of the community structure actually helps the model to predict the influence results better. The proposed community-based CNN indeed extracts the local relations of individuals. Moreover, the performance of SIL-CCNN 2 layers is better than the SIL-CCNN 1 layer. Using more hidden layers also conducts a better result.

In the real twitter dataset, the results are shown in Fig. 4. SIL-DNN and SIL-CCNN still outperform the other three methods except for the DNN_1layer. We have examined the propagation results in these two datasets. We found that the influenced scale of the synthetic dataset is much larger than the Twitter dataset. This indicates that the diffusion results in synthetic dataset should be more difficult to predict. The superiority of SIL-CCNN over SIL-DNN shows that the performance can be improved by including the community information by the proposed community-based convolutional neural network.

experts-follow-twitter/

5. Conclusions and Future Works

In this work, we proposed two neural network architectures, SIL-DNN and SIL-CCNN, to identify social influences based on the propagation of information in a social network. The proposed framework makes it possible to obtain the diffusion results within the community. SIL-DNN and SIL-CCNN can predict the users who are actually influenced from the information without the Monte Carlo simulation. Experimental results demonstrate that SIL-DNN and SIL-CCNN both outperform existing methods.

For further improvement, we will design a strategy to extend the depth of SIL-CCNN in the future to overcome the problem with the insufficient number of traces. We also want to revise the structure of neural network to consider more features such as the content of items.

References

[1] J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec. Can cascades be predicted? In *In Proceedings of WWW*, pages 925–936, 2014.

[2] C.-K. Chou and M.-S. Chen. Multiple factors-aware diffusion in social networks. In *In Proceedings of PAKDD*, pages 70–81, 2015.

[3] D. Kempe, J. M. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *In Proceedings of h ACM SIGKDD*, pages 137–146, 2003.

[4] D. Kingma and J. B. Adam. A method for stochastic optimization. In *In Proceedings of ICLR*, 2015.

[5] A. Lancichinetti and S. Fortunato. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Physical Review E*, 80, 2009.

[6] K. Saito, R. Nakano, and M. Kimura. Prediction of information diffusion probabilities for independent cascade model. In *In Proceedings of KES*, pages 67–75, 2008.

[7] O. Tsur and A. Rappoport. What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In *In Proceedings of WSDM*, pages 643–652, 2012.

[8] L. Wang, S. Ermon, and J. E. Hopcroft. Feature-enhanced probabilistic models for diffusion network inference. In *In Proceedings of ECML-PKDD*, pages 499–514, 2012.

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

A Community Sensing Approach for User Identity Linkage

Zexuan Wang Teruaki Hayashi Yukio Ohsawa

Department of Systems Innovation, School of Engineering, The University of Tokyo

User Identity Linkage aims to detect the same individual or entity across different Online Social Networks, which is a crucial step for information diffusion among isolated networks. While many pair-wise user linking methods have been proposed on this important topic, the community information naturally exists in the network is often discarded. In this paper, we proposed a novel embedding-based approach that considers both individual similarity and community similarity by jointly optimize them in a single loss function. Experiments on real dataset obtained from Foursquare and Twitter illustrate that proposed method outperforms other commonly used baselines that only consider the individual similarity.

1. Introduction

In recent years, Online Social Networks (OSNs) such as Twitter, Facebook and Foursquare tend to become the central platform of people's social life. Tons of contextual (e.g. tweets, photos) and network structure related (e.g. users' profiles, relations) data is created every day on these OSNs, which is an important resource for many valuable applications such as user behavior prediction, and cross-domain recommendation. All such applications require a crucial step called User Identity Linkage (UIL) [Shu 17], which aims to identify and link the same person/entity across different OSNs. These linkages are also called anchor links as they help align different networks under the common scene that users usually don't explicitly claim the ownership of their different accounts, and due to privacy protection rules, personal information is always restricted inside each isolated OSNs.

Abundant literature has been focusing on the UIL problem, and the majority of them fall into two categories: (1) Structure-based approaches: these approaches focus directly on the structural features of a social network, such as user names, following relationship and common neighbors between different users [Malhotra 12, Kong 13], while the problem of those approaches lies on the difficulty to find an optimal distance function between nodes to evaluate their similarity as networks are not presented in the Euclidean space [Zhang 18]. (2) Embedding-based approaches: network embedding is a new way of network representation that is able to encode the network in a continuous low-dimensional vector space while effectively preserving the network structure, for example, [Zhou 18] proposed a dual-learning embedding paradigm to improve the linking result.

However, existing methods haven't paid enough attention to the social communities naturally formed by people in the real world. Users who have limited profile information could be evaluated easier when they are located in interest groups together with their close neighbors. To better resolve the UIL problem, we proposed a novel method called Community Sensing User Identity Linkage (CSUIL), which takes advantage of both structural and embedded features of a network by designing a jointly learning model. It aids user mapping by driving some of users to the same communities they belong to, which enhances the method's accuracy and generalization ability. Experiment results on real-world dataset show feasibility of our method.

Contact: Zexuan Wang, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan, wangzexuan@g.ecc.u-tokyo.ac.jp

2. Problem Definition

Definition 1 ***Social Network Graph*** *An unweighted and undirected network is denoted as $G = \{V, E\}$, where V is the set of nodes and each node represents a user, E is the set of edges reflecting connections between nodes.*

Definition 2 ***Node Embedding*** *In a given network $G = \{V, E\}$, node embedding (a sub-task of network embedding) learns a projection function $\psi : V \mapsto \mathbb{R}^{|V| \times d}$, where $d \ll |V|$. For each node $v_i \in V$, $\boldsymbol{\psi}(v_i) \in \mathbb{R}^d$ denotes its latent representation in the vector space.*

Definition 3 ***n-th order neighbors*** *The collection of all nodes which can be reached from the given root node $v_r \in G$ within exactly n hops, denoted as $C_r = \{v_i | hop(v_i, v_r) = n\}$.*

Definition 4 ***User Identity Linkage*** *Given two different networks, $G^S = \{V^S, E^S\}$ and $G^T = \{V^T, E^T\}$. The goal of User Identity Linkage (UIL) is to predict a pair-wise linkage between a user node v_s selected from the source network G^S and an unlabeled user node v_s in the target network G^T, which indicates the same user/entity (i.e., $v^s = v^t$).*

3. Community Sensing User Identity Linkage

This proposed method consists of three main components: network embedding, community clustering and latent space mapping. A brief overview is shown in Figure 1, where blocks are the core elements in each phase, green lines indicate structural information flow directions and blue lines show how algorithms connect different phases.

3.1 Network Embedding

The quality of the latent representation of each node in both source and target network is important to the results

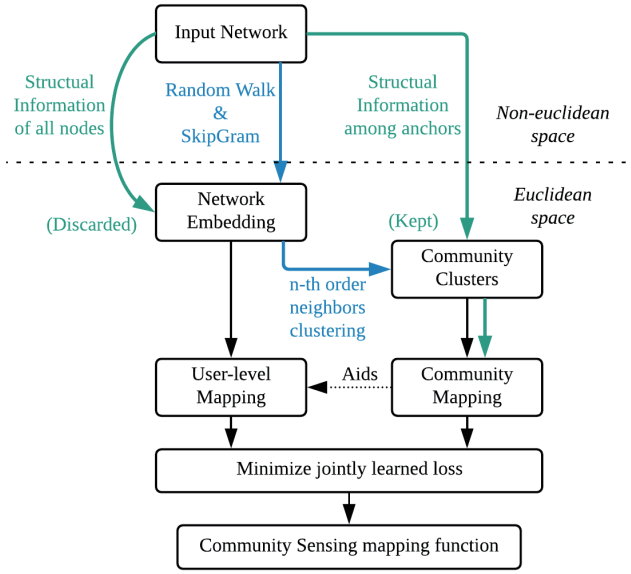


Figure 1: A brief overview of CSUIL

of the following clustering and mapping stages. Ideally, user nodes that have stronger connection, like sharing more common neighbors, or having shorter path between them should be closer to each other after they are projected into the latent space. To obtain the network embedding in good quality, an efficient model called DeepWalk [Perozzi 14] was adopted. DeepWalk mainly utilizes the truncated random walk and the SkipGram [Mikolov 13] model.

In particular, A random walk generator is first applied to the network, which will sample uniformly a random node $v_i \in G$ as the root of a random walk sequence W_{v_i}, then the generator samples uniformly from the neighbors of the last node visited until the maximum sequence length(l) is reached. The generated sequences could be thought of as short sentences, while the nodes within sequences are treated as words of a special kind of language. We could then obtain the embedding of nodes as a byproduct when updating the weight matrix in the derived SkipGram model, which aims to maximize the co-occurrence probability of nodes that appear within a window size w near the center v_j in the sequence W_{v_i}, that is to maximize the following log probability:

$$\max \quad \frac{1}{l}\sum_{i=1}^{l}\sum_{j=-w,j\neq 0}^{w} \log \Pr(v_{i+j}|v_i) \tag{1}$$

where $\Pr(v_{i+j}|v_i)$ is calculated with a hierarchical softmax function:

$$\Pr(v_{i+j}|v_i) = \frac{\exp(\boldsymbol{\psi}(v_{i+j})^{\mathrm{T}}\boldsymbol{\psi}(v_i))}{\sum_{m=1}^{l}\exp(\boldsymbol{\psi}(v_m)^{\mathrm{T}}\boldsymbol{\psi}(v_i))} \tag{2}$$

where $\boldsymbol{\psi}(v_i)$ is the embedding of node v_i we want to update at each training step and finally output to the next phase.

3.2 Community Clustering

In some supervised User Identity Linkage models such as PALE [Man 16], it only focuses on learning the user level, pair-wise matching patterns between source and target network. However, these methods failed to consider the social communities naturally formed by people in the real world. Some drawbacks may exist under such settings that users with very limited profile information could be hard to distinguish from others and the model may fall into over-fitting of local pair-wise features when trained with small amount of labeled data. More importantly, the knowledge contained in the structural relationship among anchor and non-anchor users in the original non-euclidean space is discarded after SkipGram is applied (shown by green lines in Figure 1) and later phases are not able to reuse such information.

Therefore, we made an assumption that compared to only considering the generated embedding or user-level similarity matching, the fact that which neighbors a user has in the original network, and which community a user belongs to could reveal more diffusible structural knowledge. Thus, we consider clustering the $n-th$ order neighbors of an anchor user to form their social community, the users in the same community have a closer relationship and higher similarity, which could be evaluated in some metrics including: the amount of common neighbors, or the minimum walk length between each other.

To utilize all the user information in a community, we reuse the structural information in the original network and derive a new embedding to represent this community by adopting the mean value of all community member embedding generated in Section 3.1 that are non-anchor nodes. The center that represents a certain community cluster C_i is denoted as μ_i:

$$\psi(\mu_i) = \frac{\psi(v_r) + \sum_{v' \in C_i}\psi(v')}{N+1} \tag{3}$$

where v_r is the root user, and N is the community size.

3.3 Latent Space Mapping

Let $\boldsymbol{z^s} = \psi(v^s)$ and $\boldsymbol{z^t} = \psi(v^t)$ be the node embedding generated in Section 3.1 and the final stage of CSUIL is Latent Space Mapping. In this phase, we try to find a mapping function from the source network to the target network Φ: $\mathbb{R}^{|V^s|\times d} \mapsto \mathbb{R}^{|V^t|\times d}$, that will minimize the distance between the predicted embedding $\Phi(\boldsymbol{z^s})$ and the true corresponding embedding $\boldsymbol{z^t}$ of $\boldsymbol{z^s}$ in the target network:

$$\min \quad \left\|\Phi\left(\boldsymbol{z^s}\right) - \boldsymbol{z^t}\right\|_F \tag{4}$$

We then train a novel two-inputs and two-outputs neural network model, which breaks down the whole task above into two simultaneously conducted parts: (1) minimize the distance between predicted and real user node (2) minimize the distance between predicted and real community center. The second sub-task will drive the mapping function to the direction that also exploits the relationship between community centers in both source and target networks to increase the generalization ability of the model on new unseen data.

Next, the design of the loss function could be one of the most critical parts of a machine learning model, a good loss function should reflect the error during training as well as the generalization error that guides parameters to optimize the model. Therefore, for the goal of above two sub-tasks, a new community sensing loss function is proposed as:

$$loss = (1-\gamma) \sum_{(v^s,v^t)\in\{S,T\}} \left\|\Phi(\boldsymbol{z^s};\theta) - \boldsymbol{z^t}\right\|_F + \gamma \sum_{\mu\in C} \left\|\Phi(\boldsymbol{\mu^s};\theta) - \boldsymbol{\mu}^t\right\|_F \quad (5)$$

where $\{S,T\}$ is the set of groundtruth anchor pairs, C is the set of community centers, F is the Frobenius norm, θ is the collection of all parameters in the model, and γ is the hyper-parameter of the weight coefficient of the community loss that could be co-optimized during the learning of the mapping function.

We finally employed a Multi-Layer Perceptron (MLP) model that does not require extensive feature selection or difficult parameter tuning to learn the optimized mapping function, while this model also has the flexibility of dealing with the non-linear relationships that may exist between the source and target network.

The whole algorithm design is shown in Algorithm 1.

Algorithm 1: CSUIL

Input: network $G(V,E)$, anchor nodes $\{S,T\}$, test nodes $\{S',T'\}$, community clustering parameter n, community loss parameter γ
Output: mapping function Φ, matching result list R
foreach *node* $v_i \in G$ **do**
  Generate the embedding of v_i as $\boldsymbol{z_i}$
end
foreach *anchor node pair* $\{s_i,t_i\}$ *in* $\{S,T\}$ **do**
  Reuse the original network structure information, cluster the n-order neighbors of s_i and t_i
  Derive the community center $\boldsymbol{\mu_i^s}$ and $\boldsymbol{\mu_i^t}$
end
Train the MLP model by jointly minimize the node mapping loss $\left\|\Phi(\boldsymbol{z^s};\theta) - \Phi(\boldsymbol{z^t})\right\|_F$ and community loss $\left\|\Phi(\boldsymbol{\mu^s};\theta) - \Phi(\boldsymbol{\mu^t})\right\|_F$
foreach *test node* $s'_i \in S'$ **do**
  Add the predicted t'_i to result list R
end
Evaluate(R,T')

4. Experiment

4.1 Data Preparation

A real-world social network dataset collected from Twitter and Foursquare [Zhang 15] is used in this experiment, which was released in [Liu 16]. All the sensitive personal information is removed under privacy concerns to form the final training and testing data. The ground truth of anchors is obtained by crawling users' Twitter accounts from their Foursquare homepage. Table 1 lists the statistics of this dataset.

Network	#Users	#Relations	#Anchors
Twitter	5,220	164,919	1,609
Foursquare	5,315	76,972	

Table 1: Statistics of Twitter-Foursquare Dataset

4.2 Evaluation Metrics

In this experiment, in a similar form to [Zhou 18], a metric called *Precision*@k was adopted, which is defined as:

$$Precision@k = \frac{\sum_{i}^{n} TOP_k(\Phi(\boldsymbol{z_i^s}))}{N} \quad (6)$$

where $TOP_k(\Phi(\boldsymbol{z_i^s}))$ is a binary output function (0 or 1), for each predicted embedding $\Phi(\boldsymbol{z_i^s})$, it tells whether the positive match $\boldsymbol{z_i^t}$ exists in the $top-k$ list or not, and N is the number of all testing nodes. In the context of UIL, as *Precision*@k is a metric of the true positive rate, it could be treated the same as *Recall*@k, and F_1@k.

4.3 Comparative Methods

We compare the proposed CSUIL with several existing embedding-based methods, and take them as the baseline of this task.

- **CSUIL**: the proposed method, it could explicitly exploit the individual as well as community features of a network, by jointly optimizing mapping functions that concentrate on user-level and community-level similarity respectively.
- **IONE**: Proposed in [Liu 16] and adopted as a baseline result, Input-Output Network Embedding (IONE) is a network embedding and partial network alignment method. It takes follower-ship and followee-ship as input and output contexts and generates all three representations together with the user node.
- **INE**: INE is a simplified version of IONE, which only consider node and input representation for matching.

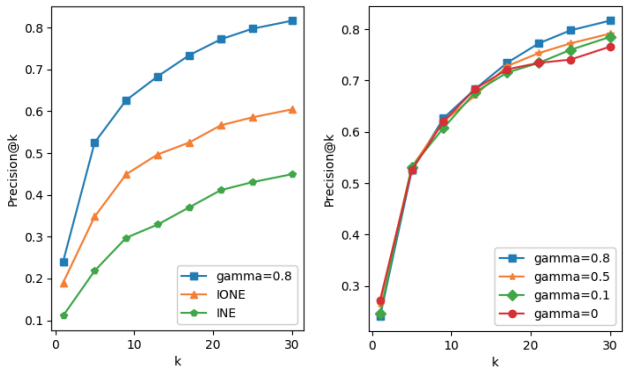
4.4 Results

The performance results are illustrated in Table 2 and Figure 2. In the experiment, during the community clustering phase, the cluster size is set to first-order neighbors for the simplicity. Then we examine the ability of the final model (with training rate=90%) on the link prediction task. For CSUIL, we report the result in different settings of precision metrics k and community loss weight coefficient γ. For INE and INOE, we report the result in the original paper's default setting.

	Precision@k							
γ	P@1	P@5	P@9	P@13	P@17	P@21	P@25	P@30
INE	0.1108	0.2184	0.2975	0.3291	0.3703	0.4114	0.4304	0.4494
IONE	0.1899	0.3481	0.4494	0.4968	0.5253	0.5665	0.5854	0.6044
0.8	**0.2405**	**0.5190**	**0.6203**	**0.6835**	**0.7342**	**0.7722**	**0.7975**	**0.8165**

Table 2: Performance comparison between baselines

From the experiment results, we could conclude that:



(a) Baselines (b) Precision on different γ

Figure 2: Link prediction precision results. X-axis is the different value of k, for the top-k list being evaluated; Y-axis is the precision result in percentage.

- Compared to the baseline model, INE and IONE, the best performance (when $\gamma = 0.8$, shown in Table 2 and Figure 2(a)) of our approach has an improvement from about 6% to 21 % at most in different settings of precision metrics, which shows the feasibility of this approach.
- Figure 2(b) also illustrates that by changing the setting of community loss weight coefficient γ, the ability of the model to sense more positive matching in a larger search space (higher k setting in precision), could be enhanced, which is an important improvement because many other papers only stress their performance at the $k = 30$ setting. However, adding too much weight to community loss may lead to a slight reduction of the ability to narrow the target to a finer scale (lower k in precision), compared with the $\gamma = 0$ setting.

5. Conclusion

In this paper, we aim to study the UIL problem by reusing the discarded knowledge in the original Online Social Network after network embedding. Not limited to anchor same users across networks, we would also like the community formed by close users to have a positive match across networks. This is because some users may have limited profile and it could be hard to distinguish them from others. However, in the context of a community, users share common features, and they will be driven to the correct direction where group of users with high similarity locates, even if community members are known little. This could also help to avoid overfitting the input data and increase the generalization ability of the method.

Therefore, we break down the main task into two simultaneously learned sub-tasks: User Mapping and Community Mapping, this is achieved by jointly optimizing the user loss and community loss in a single MLP model. Based on above theories, Community Sensing User Identity Linkage (CSUIL) is proposed. Results show that our approach outperforms current baseline models, and has the flexibility to adapt hyper-parameters for different needs or data input.

Acknowledgments

This work was funded by JSPS KAKENHI JP16H01836, JP16K12428, and industrial collaborators.

References

[Kong 13] Kong, X., Zhang, J., and Yu, P. S.: Inferring anchor links across multiple heterogeneous social networks, in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pp. 179–188ACM (2013)

[Liu 16] Liu, L., Cheung, W. K., Li, X., and Liao, L.: Aligning Users across Social Networks Using Network Embedding., in *IJCAI*, pp. 1774–1780 (2016)

[Malhotra 12] Malhotra, A., Totti, L., Meira Jr, W., Kumaraguru, P., and Almeida, V.: Studying user footprints in different online social networks, in *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, pp. 1065–1070IEEE Computer Society (2012)

[Man 16] Man, T., Shen, H., Liu, S., Jin, X., and Cheng, X.: Predict Anchor Links across Social Networks via an Embedding Approach., in *IJCAI*, Vol. 16, pp. 1823–1829 (2016)

[Mikolov 13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J.: Distributed representations of words and phrases and their compositionality, in *Advances in neural information processing systems*, pp. 3111–3119 (2013)

[Perozzi 14] Perozzi, B., Al-Rfou, R., and Skiena, S.: Deepwalk: Online learning of social representations, in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 701–710ACM (2014)

[Shu 17] Shu, K., Wang, S., Tang, J., Zafarani, R., and Liu, H.: User identity linkage across online social networks: A review, *Acm Sigkdd Explorations Newsletter*, Vol. 18, No. 2, pp. 5–17 (2017)

[Zhang 15] Zhang, J. and Philip, S. Y.: Integrated Anchor and Social Link Predictions across Social Networks., in *IJCAI*, pp. 2125–2132 (2015)

[Zhang 18] Zhang, J.: Social Network Fusion and Mining: A Survey, *CoRR*, Vol. abs/1804.09874, (2018)

[Zhou 18] Zhou, F., Liu, L., Zhang, K., Trajcevski, G., Wu, J., and Zhong, T.: DeepLink: A Deep Learning Approach for User Identity Linkage, in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pp. 1313–1321IEEE (2018)

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Learning Sequential Behavior for Next-Item Prediction

Na Lu Yukio Ohsawa Teruaki Hayashi

Department of System Innovation, School of Engineering, The University of Tokyo

A more precise recommendation plays an essential role in e-commerce. Representation learning has attracted many attentions in recommendation field for describing local item relationships. In this paper, we utilize the item embedding method to learn item representations and user representations. Our methods compute cosine similarity of user vector and recommended item vectors to achieve the goal of personalized ranking. Experiment on real-world dataset shows that our model outperforms baseline model especially when the number of the recommended item is relatively small.

1. INTRODUCTION

The sharp growth of e-commerce and the using mobile electronic device require a more precise prediction of next item that users would probably like to purchase. Data mining of users' behaviors aims at finding useful patterns from a large database. In this task, understanding users' history and features are one of the most critical parts.

To deal with this task, some models were developed based on last transaction information, which is mostly involving Markov chains[Chen 12]. This method mainly makes use of users' sequential transaction data to predict what will be the next item considering the last transaction event. The advantages of this method are that it could consider the time sequence and recommend a proper item for the next movement. Other general recommendation models would consider users' past purchase behavior as a whole to generate their overall taste (or features)[Rendle 10] . This method could generally grasp a user's interests. The most widely used method of general recommendation models is called collaborative filtering. The advantages of this method are that it could get users' interesting points. Thus, the recommendation could generate from users' whole behavior. However, this method discards subsequent information that may lack preciseness in next-item prediction.

Here a good recommendation model could consider not only the sequential information but also users' overall taste. A hierarchical representation model was proposed to combine both sequential information and user history transaction information [Wang 15]. The proposed hierarchical representation model used a two-layer model. One-layer aggregated all the sequential transactions, and in the second layer, this sequential information was aggregated with the user's overall taste. Then the combined information was used to predict item in the next transaction. This method was novel by setting different layers to combine two kinds of information. However, a better method has been proposed to learn item representations.

Contact: Na Lu, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 Japan, Department of Systems Innovation, School of Engineering,The University of Tokyo, Bldg.No.8. 507 , 080-1241-0956, luna@g.ecc.u-tokyo.ac.jp

For understanding sequence data, we utilize the Skip-Gram model for word representation learning in natural language process field [Mikolov 13] named as word2vec. Skip-Gram model learns word representations by predicting the context of this word. More precisely, word2vec get a word vector in a lower dimensional space compared with one-hot representation. This method was later generalized as item2vec for learning item representations [Barkan 16]. Item2vec treats users' subsequent behavior as a sentence in word2vec and creates item vectors.

By learning users' sequential data to generate item representations, we proposed a method for aggregating users' history behavior and general taste to build a recommendation system.

2. RELATED WORKS

A good recommendation system could improve users' decision-making process in this information overload era. The widely used recommendation methods include collaborative filtering, content-based filtering, and hybrid filtering. Despite the use of traditional methods, many approaches are proposed to improve the quality of recommendations. We first review some related work in this field.

2.1 Sequential Pattern Mining

Pattern mining is an essential branch of data mining, which consists of discovering frequent itemsets, associations, sub-graphs, sequential rules, etc. [Chen 96]. The target of sequential pattern mining is to detect sequential patterns by analyzing a set of sequential data, in which occurrence frequency is one of the target [Pei 04]. Item2vec embeds items into a low-dimensional representation by accounting the item co-occurrence in user records. That is, this model could generally capture the co-occurrence patterns of items in each transaction data.

2.2 Personalized Ranking

From the target of the recommendation system, it can be treated as a rating prediction problem or a personalized ranking problem [Rendle 09]. The task of personalized ranking is to provide a user with a ranked list of items, which matches a real-life scenario. An example is that an online retailer wants to give a personalized ranking item list that a user may probably buy in the recent future. For-

mer research for personalized ranking algorithms optimized through learning users' preferences on a set of items, which include BPR [Rendle 09], CliMF [Shi 12].

2.3 Item Representation Learning

The word embedding method [Mikolov 13] have attracted much attention from fields besides NLP. The recommendation is also to utilize this method for better performance, including clustering [Barkan 16] and regression. Representation learning in recommendation means getting relationships between items from a specific data set, which is called item embedding. Barkan and Koenigstein [Barkan 16] first proposed Item2Vec model which based on a neural item embedding model for collaborative filtering. In this method, item embedding is used to learn a better item representation but fail to give a personalized ranking recommendation. In this research, we propose an item embedding based method combined with users' history behaviors to provide a personalized next-item recommendation.

3. PROBLEM STATEMENT

In this section, we first introduce the problem formalization of recommendation based sequence behavior. We then describe the item embedding and recommendation for the next item in detail. After that, we talk about the learning and prediction procedure of this method.

3.1 Formalization

Let $U = \{u_1, u_2, ..., u_{|U|}\}$ be a set of users and $I = \{i_1, i_2, ..., i_{|I|}\}$ be a set of items, in which $|U|$ and $|I|$ denote to the total number of unique users and items, respectively. For each user u, the transaction history data T^u is given by $T^u = (T_1^u, T_2^u, T_3^u, ..., T_t^u)$, where $T_t^u \subseteq I$. The purchase history of all users is denoted as $T = \{T^1, T^2, T^3, ..., T^t\}$. Given the transaction data of all users, our task is to predict what the user will probably buy in the next time (eg. t-th), which is denoted as $R = \{R^1, R^2, R^3, ..., R^u\}$. Every R^i includes k items as recommendation: $R^i = \{R_1^i, R_2^i, ..., R_k^i, \}$.That is, we need to generate a personalized ranking R^i for user u_i in t-th transaction.

3.2 Item2Vec algorithm

Our purpose is to learn a recommendation model from a sequential transaction data which could also combine users' overall taste. In this section, we first explain Item2Vec algorithms in detail, which generate item embedding from sequential data. Then users' general taste will be concluded from one user's whole transaction data. At last, item representations and users' general taste will be combined to create a personalized ranking for a next-item recommendation.

To proposed our method for personalized ranking from a sequential user transaction data, we first need to have a look at Item2Vec specifically. Skip-gram with negative sampling (SGNS) was first introduced in word embedding by Mikolov et al. [Mikolov 13]. The neural embedding in natural language processing attempts to map words and phrases into a vector space of low-dimensional semantics and syntax. Skip-gram uses the current word to predict its context words. The item collection in Item2vec is equivalent to the sequence of words in word2vec, that is, the sentence. Commodity pairs that appear in the same collection are considered positive. For the set $w_1, w_2, ..., w_K$ objective function:

$$\frac{1}{K}\sum_{i=1}^{K}\sum_{j\neq i}^{K} log(w_j|w_i) \tag{1}$$

Same as word2vec, using negative sampling, define $p(w_j|w_j)$ as:

$$p(w_j|w_j) = \sigma(u_i^T v_j)\prod \sigma(-u_i^T v_k) \tag{2}$$

Finally, the SGD method is used to learn the max of the objective function and to obtain the embedding representation of each item. The cosine similarity between the two items is the similarity of items.

The cosine similarity between two vectors can be formalized as:

$$cos(v_1, v_2) = \frac{v_1 \cdot v_2}{|v_1||v_2|} \tag{3}$$

3.3 Proposed method

From Item2Vec method, all users' transaction data $T = \{T^1, T^2, T^3, ..., T^t\}$ is used to learn item representations. More specifically, Item2Vec algorithm inputs a large corpus of transactions and creates a vector space, in which every unique item is transformed as a vector in this space. Based on this, we produce item representations based on users' sequential transaction data.

The advantage of our methods is that we can introduce aggregation operations in forming user representations from their history transaction data. In this work, we propose two aggregation methods to get a user representation.

The first is average pooling. This method construct one vector by taking the average value from a set of vectors. Let $V = \{v_1, v_2, v_3, ..., v_l\}$ be a set of vectors. Average pooling of V can be formalized as:

$$f_{ave}(V) = \frac{1}{l}\sum_{i=1}^{l} v_i \tag{4}$$

Second is max pooling. This method construct one vector by taking the max value from a set of vectors. Thus, max pooling can be formalized as:

$$f_{max}(V) = \begin{bmatrix} max(v_1[1] & ... & v_l[1]) \\ max(v_1[2] & ... & v_l[2]) \\ ... & ... & ... \\ max(v_1[n] & ... & v_l[n]) \end{bmatrix} \tag{5}$$

From a user's transaction data T^i, we can get a user representation $\vec{u_i}$ from $f_{ave}(T^i)$ and $f_{max}(T^i)$ as $u_{i\vec{ave}}$ and $u_{i\vec{max}}$. Combine with top-K recommendation from item embedding, which is R^i, we re-rank R^i based on the weighted similarity with user u_i. The detail of re-ranking of recommendation R^i is in Algorithm 1.

In this way, we can combine u_i's general taste ($u_{i\vec{ave}}$ and $u_{i\vec{max}}$) and sequential prediction (R^i) to get a overall prediction.

Algorithm 1 Combination of user representation and top-K recommendation

Input: top-K recommendation R^i for u_i, user vector $u_{i\vec{ave}}$ and $u_{i\vec{max}}$, item set I
Output: R^i_{ave} and R^i_{max}

```
for j ≤ top − K ∗ 2 do
    if R^i_j ⊆ I then
        R^i_{ave−j} = cos(R^i_j, u_{iave}) and R^i_{max−j} = cos(R^i_j, u_{imax})
    else
        test_ size − 1
    end if
end for
sort R^i_ave and R^i_max from highest to lowest, choose top-K items from R^i_ave and R^i_max
return R^i_ave and R^i_max
```

Dataset name	# users	# items	# T
Online Retail	90,346	2553	397,923

Table 1: Basic Information about Online Retail dataset

4. EXPERIMENT AND DISCUSSION

In this section, we conduct empirical experiments to test the effectiveness of our method for a next-item recommendation. We first introduce the experimental data set, the baseline methods in our experiments. Then we compare our approach with the baseline model to study the effect of different aggregations. Finally, we make some analysis on the result of the experiments.

4.1 Dataset

We conduct our experiment on an open data set named 'Online Retail dataset' [UCI 15]. This data set includes transaction data from 2010.12.01 to 2011.12.09. Every row includes invoice number, product number, product name, sale quantity, sale time, unit price, customer ID, and customer's country. After deleting the row that has a default value, the data set basic information is in Table 1.

4.2 Evaluation and Discussion

We divided the dataset into train data and test data. Train data was used to train item2vec model to generate the item representations. Test data was used to evaluate the effectiveness of our method.

In the test data, we first remove the last transaction data from user u. So the remaining is $T^u_{n-1} = \{i_1, i_2, i_3, ..., i_{t-1}\}$. We use the learned model and remaining T^u to make a recommendation of top-K items located closer to each item in the learned vector space. Then these top-K items and user vector derived from T^u_{n-1} are combined to make the final top-K recommendation.

Here we use Recall as the prediction evaluation. The recall is formalized as below:

$$Recall(T^u_t, R^u_t) = \frac{T^u_t \bigcap R^u_t}{T^u_t}$$

top-K \ Method	Ave	Max
1	14.98%	20.25%
3	6.41%	8.43%
5	8.95%	7.54%
10	4.57%	6.13%
15	1.47%	7.75%
20	3.37%	3.85%

Table 2: Recall percentage improvement compared with baseline method

In our experiment, we set top-K=1,3,5,10,15,20 as the number of items that would be recommended to user u. In this experiment, the baseline method is the prediction derived from the item2vec method, which was not combined with a user vector. The comparison of these methods are as follows.

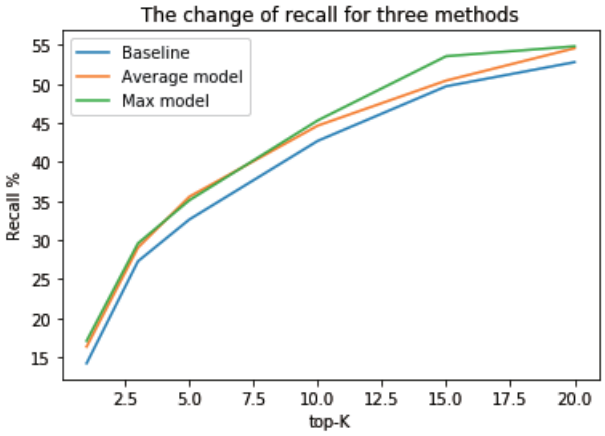


Figure 1: The change of recall for three methods

We can see that the average model and max model could improve over 10% of recall compared with the baseline model. That means if we recommend one item for a user, our model performed well by aggregating user's vector and item2vec prediction. However, this improvement declined with the increase in top-K, which means if we recommend a lot of items to a user at one time, our improvement is not as effective as recommending fewer items. Compared with the baseline model, the recall of the average model and max model are higher, and they get higher with the increase of top-K. If we provide more items for a user, the probability of correct prediction will be higher, just as Figure 1 shows above.

5. CONCLUSION

Representation learning has attracted many attentions in recommendation field for describing local item relationships. In this paper, we utilize the item embedding method to learn item representations from sequential transaction data. And we also constructed user representations to get a ranked list of items for a user. The experiment result

demonstrated that our proposed method for next-item recommendation outperformed baseline model in prediction recall. Specifically, our models get 14.98% and 20.25% improvement compared with baseline model in a top-1 recommendation, which means we get a distinct improvement when the number of the recommended item is relatively small.

6. ACKNOWLEDGEMENT

This work was funded by JSPS KAKENHI, JP16H01836, JP16K12428, and industrial collaborators.

References

[Chen 12] Chen, Shuo, et al. "Playlist prediction via metric embedding." Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012.

[Rendle 10] Rendle, Steffen, Christoph Freudenthaler, and Lars Schmidt-Thieme. "Factorizing personalized markov chains for next-basket recommendation." Proceedings of the 19th international conference on World wide web. ACM, 2010.

[Wang 15] Wang, Pengfei, et al. "Learning hierarchical representation model for nextbasket recommendation." Proceedings of the 38th International ACM SIGIR conference on Research and Development in Information Retrieval. ACM, 2015.

[Mikolov 13] Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." Advances in neural information processing systems. 2013.

[Barkan 16] Barkan, Oren, and Noam Koenigstein. "Item2vec: neural item embedding for collaborative filtering." Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on. IEEE, 2016.

[Chen 96] Chen, Ming-Syan, Jiawei Han, and Philip S. Yu. "Data mining: an overview from a database perspective." IEEE Transactions on Knowledge and data Engineering 8.6 (1996): 866-883.

[Pei 04] Pei J, Han J, Mortazavi-Asl B, et al. Mining sequential patterns by pattern-growth: The prefixspan approach[J]. IEEE Transactions on Knowledge & Data Engineering, 2004 (11): 1424-1440.

[Rendle 09] Rendle, Steffen, et al. "BPR: Bayesian personalized ranking from implicit feedback." Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence. AUAI Press, 2009.

[Shi 12] Shi, Yue, et al. "CLiMF: learning to maximize reciprocal rank with collaborative less-is-more filtering." Proceedings of the sixth ACM conference on Recommender systems. ACM, 2012.

[UCI 15] https://archive.ics.uci.edu/ml/datasets/Online+Retail

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Application of Unsupervised NMT Technique to Japanese–Chinese Machine Translation

Yuting Zhao*1 Longtu Zhang*2 Mamoru Komachi*3

Tokyo Metropolitan University

Neural machine translation (NMT) often suffers in low-resource scenarios where sufficiently large-scale parallel corpora cannot be obtained. Therefore, a recent line of unsupervised NMT models based on monolingual corpus is emerging. In this work, we perform three sets of experiments that analyze the application of unsupervised NMT model in Japanese–Chinese machine translation. We report 30.13 BLEU points for ZH–JA and 23.42 BLEU points for JA–ZH.

1. Introduction

Neural machine translation (NMT) has recently shown impressive results thanks to the availability of large-scale parallel corpora [Bahdanau 14]. NMT models typically fit hundreds of millions of parameters to learn distributed representations which may generalize better when data is redundant. Unfortunately, finding massive amounts of parallel data remains challenging for vast majority of language pairs, especially for low-resource languages, as it may be too costly to manually produce or nonexistent. Conversely, monolingual data is much easier to find, and many languages with limited parallel data still possess significant amounts of monolingual data.

Recently, remarkable results have been shown in training NMT systems relying solely on monolingual data in the source and target languages by using an unsupervised approach [Artetxe 18, Lample 18a]. They proposed unsupervised NMT models that are effective on English–French and English–German. Following their practice, we try to apply unsupervised NMT model to Japanese–Chinese translation.

In this work, we perform experiments from two data domains. They are divided into two types of monolingual corpus and quasi-monolingual corpus. Among them, the best BLEU score can reach 30.13 of ZH–JA and 23.42 of JA–ZH with using ASPEC-JC (Japanese Chinese language pairs) parallel corpus [Nakazawa 16] in the quasi-monolingual setting.

2. System Architecture

The unsupervised NMT model [Lample 18b] we used is composed of two encoder-decoder models for source and target languages and in series with back-translation models. The encoders will encode monolingual sentences into latent representations for respective decoders. One decoder is used as a translator to decode the latent representations, and the other decoder perform the denoising effect of a language model on the target side that refines the latent representation of the source sentence. Then, it jointly train two back-translation models together with the two encoder-decoder language models. In the forward translation, the model generates data which will be trained to the backward translation and in the backward translation, the model trained from the generated target to the source generates translations. The generated sentences from back-translation are added to the regular training set in order to regularize the model.

Contact: Yuting Zhao, zhao-yuting@ed.tmu.ac.jp

3. Experiment

3.1 Datasets

We prepare three data sets from ASPEC-JC (Japanese Chinese language pairs) parallel corpus [Nakazawa 16] and Wikipedia dump *1.

For quasi-monolingual data, the Japanese–Chinese portion of ASPEC-JC was used. Note that although this is a parallel corpus, we shuffled it and used it monolingually. In this paper, we call it ASPEC-Quasi. Official training/development/testing split contains totally 670,000 Chinese and Japanese sentences for training and 2,000+ sentences for evaluating and testing.

For monolingual data, the Japanese–Chinese portion of ASPEC-JC was also used. Note that we shuffled it monolingually and divided the monolingual Chinese and Japanese data into the first half and the second half. Then, they were staggered and combined to form two groups of monolingual data sets with a size of 335,000, and one group was randomly selected for experiment. In this paper, we call it ASPEC-Mono. In addition, we created a Japanese–Chinese monolingual corpus with a training size of 10 million from Wikipedia articles. As above, evaluation and test data are all official data from ASPEC-JC.

3.2 Preprocessing

Firstly, we tokenize Japanese and Chinese datasets separately. We use MeCab *2 with dictionary IPADic for Japanese and Jieba *3 with its default dictionary for Chinese. Secondly, we join the source and target monolingual corpora to learn fastBPE tokens with the vocabulary size of 30,000. Finally, we apply fastText [Bojanowski 17] on the

∗1 https://dumps.wikimedia.org/
∗2 http://taku910.github.io/mecab/
∗3 https://github.com/fxsjy/jieba

Corpora	Amount	JA-ZH	ZH-JA
ASPEC-Mono	335,000	8.9	10.37
Wikipedia	10,000,000	9.74	12.51
ASPEC-Quasi	670,000	**23.42** *(31.19)*	**30.13** *(39.18)*

Table 1: BLEU scores of 3 datasets. (The BLEU score of OpenNMT model is presented in parentheses)

BPE tokens. This way, we obtain cross-lingual BPE embeddings for Chinese and Japanese language pairs to initialize lookup tables. More specifically, we use the skip-gram model with ten negative samples, a context window of 5 words, and 512 dimensions.

3.3 Model

In this work, our models use transformer cells as basic units in the encoders and decoders with PyTorch toolkit version 0.5. We set the number of layers of both the encoders and decoders to 4, and the hidden layers is set to 512. Adam optimizer is used with a learning rate of 0.0001 and a batch size of 25. We set a maximum length of 175 tokens per sentence for each type of dataset and a dropout rate of 0.1. We also set random blank-out rate to 0.1 and word shuffle of 3.

BLEU score is used to evaluate translation in both directions with every iteration, and training will stop when the scores from the last 3 iteration did not improve any more.

4. Results and Discussions

The BLEU scores obtained by all the tested datasets are reported in Table 1.

Amount of data. Firstly, we see the results obtained from the complete monolingual datasets ASPEC-Mono and Wikipedia. As our baseline, the results of ASPEC-Mono obtained 8.9 BLEU points for JA–ZH and 10.37 BLEU points for ZH–JA. As the amount of sentences increases from 335,000 to 10,000,000, the results of Wikipedia obtained 9.74 BLEU points for JA–ZH and 12.51 BLEU points for ZH–JA. Comparing with ASPEC-Mono, scores have gone up in both directions despite of domain difference.

Quasi-monolinguality. Secondly, we see the results in the last row, which is from ASPEC-Quasi corpus. It gets 23.42 BLEU points for JA–ZH and 30.13 BLEU points for ZH–JA. This result exceeds all the previous two results. Moreover, the OpenNMT model using ASPEC-JC parallel corpus reports 31.19 BLEU points for JA–ZH and 39.18 BLEU points for ZH–JA. In contrast, the BLEU score of unsupervised NMT is lower than that of supervised NMT, but the gap is not big.

5. Related Work

From the work of Sennrich et al. [Sennrich 16], they proposed a straightforward approach to create synthetic parallel training data by pairing monolingual training data with an automatic back-translation.

Recently, Artetxe et al. [Artetxe 18] and Lample et al. [Lample 18a] have achieved substantial improvement for fully unsupervised machine translation. They leverage strong language models through training the sequence-to-sequence system as a denoising autoencoder.

6. Conclusion

Based on the above analysis, it can be inferred as follows:

- For monolingual data, the larger the data, the better the translation results.
- For quasi-monolingual data, the effectiveness of unsupervised NMT model on Japanese–Chinese is quite promising, even if it uses smaller training dataset.

From the experiment, we can see unsupervised NMT is effective in Japanese–Chinese machine translation. However, it is worth considering that why there is a huge gap between the results of using monolingual corpus and quasi-monolingual corpus on Japanese–Chinese unsupervised NMT. Even though the amount of monolingual Wikipedia corpus is 15 times more than that of ASPEC-Quasi corpus, the result is much worse. We hope to start from this significant gap and continue to study the factors affecting unsupervised NMT in Japanese–Chinese machine translation.

References

[Artetxe 18] Artetxe, M., Labaka, G., Agirre, E., and Cho, K.: Unsupervised Neural Machine Translation, in *Proceedings of ICLR* (2018)

[Bahdanau 14] Bahdanau, D., Cho, K., and Bengio, Y.: Neural Machine Translation by Jointly Learning to Align and Translate, in *Proceedings of ICLR* (2014)

[Bojanowski 17] Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T.: Enriching word vectors with subword information, in *Proceedings of TACL*, pp. 135–146 (2017)

[Lample 18a] Lample, G., Conneau, A., Denoyer, L., and Ranzato, M.: Unsupervised Machine Translation Using Monolingual Corpora Only, in *Proceedings of ICLR* (2018)

[Lample 18b] Lample, G., Ott, M., Conneau, A., Denoyer, L., and Ranzato, M.: Phrase-Based & Neural Unsupervised Machine Translation, in *Proceedings of EMNLP*, pp. 5039–5049 (2018)

[Nakazawa 16] Nakazawa, T., Yaguchi, M., Uchimoto, K., Utiyama, M., Sumita, E., Kurohashi, S., and Isahara, H.: ASPEC: Asian Scientific Paper Excerpt Corpus, in *Proceedings of LREC*, pp. 2204–2208 (2016)

[Sennrich 16] Sennrich, R., Haddow, B., and Birch, A.: Improving Neural Machine Translation Models with Monolingual Data, in *Proceedings of ACL*, pp. 86–96 (2016)

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Synthetic and Distribution Method of Japanese Synthesized Population for Real-Scale Social Simulations

Tadahiko Murata[*1]

[*1] Department of Informatics, Kansai University, Japan

Takuya Harada[*2]

[*2] Research Institute for Socionetwork Strategies, Kansai University, Japan

In this paper, we describe how synthesized populations are essential in real-scale social simulations (RSSS), and the current situation of the population synthesis for whole populations in Japan. RSSS is simulations using the real number of populations or households in social simulations. This paper describes how we have completed to synthesize multiple sets of populations based on the statistics of each local government in Japanese national census in 2000, 2005, 2010 and 2015. We have started to distribute those multiple sets of the synthesized populations for researchers of RSSSs in Japan. In distributing the synthesized populations, we should protect personal or private information in the synthesized populations. We show some scheme how to protect them using a cloud service or secure computations.

1. Introduction

In this paper, we try to develop a platform for Real-Scale Social Simulation (RSSS) by synthesizing whole households in Japan and providing the data of synthesized households for researchers who try to develop RSSS tools. RSSS is simulations using populations or households in the real scale.

Recently social simulations have attracted from many researchers to tackle with problems in our environments or communities. One of the most influential social simulations is the segregation model proposed by **Schelling (1971)**. In his model, he clearly shows how segregations happen due to the preference of residents to be a neighbor of the same race or group. His model shows that segregations can happen even if there is no hostility among races. His model is quite interesting and meaningful to give understanding of conflict and cooperation. He was awarded the 2005 Novel Memorial Prize in Economic Science.

Although Schelling's model is quite significant, interpretation is required to apply his model to real situations. If we are able to directly conduct simulations with real-scale environments and real-scale residents, it is easy to draw insight from simulation results. That is why RSSS has much attention from many researchers recently.

In order to conduct RSSS, real-scale populations are required. For example, when **Murata & Konishi (2013)** optimized the number of polling places with considering the voting rate and the number of polling places in a city using a scheme of RSSS, they should synthesize the population in the city and measure the distance of polling places from their homes. When **Murata & Du (2015a)** assessed effects of the pension program for each household in Japan, they should create and simulate demographic movement of all prefectures in Japan for 25 or over 100 years according to the statistics of Japanese census conducted in 2010.

2. Synthetic reconstruction methods

Tadahiko Murata; Address: Kansai University, 2-1-1, Ryozenji, Takatsuki, Osaka 569-1095, Japan; Phone: +81-72-690-2429, FAX; +81-72-690-2491; murata@kansai-u.ac.jp .

Since RSSS researchers should face to synthesize populations in the target area of their social simulation sooner or later, we have synthesized populations using the available statistics in each local government such as city, town and village in Japan according to the national census in 2000, 2005, 2010 and 2015. The number of cities, towns and villages in Japan is 1741 in the national census in 2015).

Methods synthesizing populations with individual attributes are known as Synthetic Reconstruction method (SR method) (**Wilson, 1976**). Originally an SR method employs real samples from the real statistics. That method increases the number of individuals from the samples in order to fit the real statistics. Here, we prefer using the term "synthesize" to the term "reconstruct" in this paper. Since a reconstruction method is expected to generate exactly the same attributes of each individual in the population, however, it is impossible to reconstruct the same attributes from a small number of statistics. Therefore, we can only synthesize a population that has the same statistical characteristics using SR methods. **Lenormand & Deffuan (2013)** compared SR methods that employ samples with a synthetic method without samples. They showed the synthetic method without samples is better than the former one.

We employ a synthetic method without samples in this paper. The basis of our method is a method proposed by **Ikeda et al. (2010)**. They proposed a method for synthesizing households of nine family types according to the nine real statistics using a simulated annealing method **(Davis, 1987)**. **Fig. 1** shows the nine

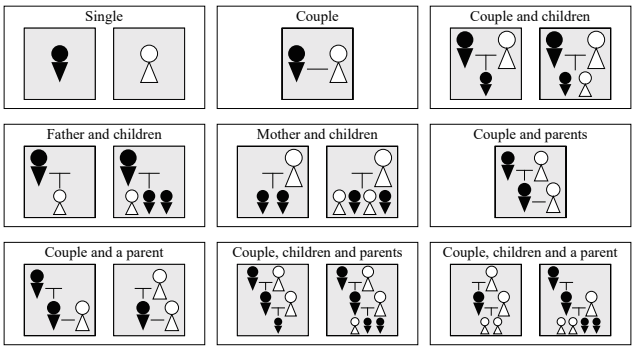
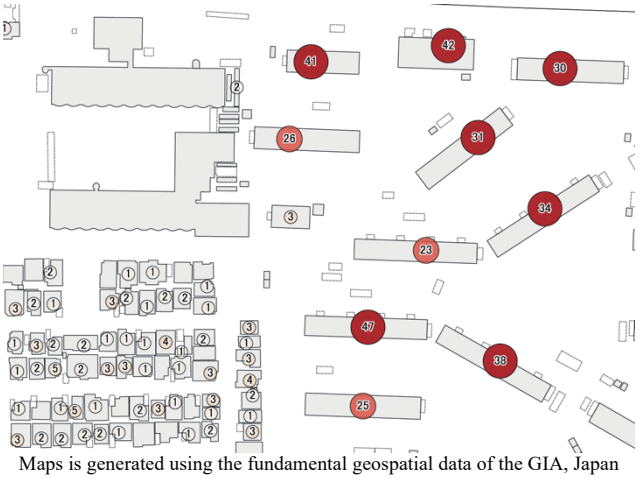


Fig. 1 Nine Family Types.



Maps is generated using the fundamental geospatial data of the GIA, Japan

Fig. 2 Households Projection on buildings.

family types they synthesized. 95% households in Japan come from these family types. Each family member has attributes of sex, age, kinship in its household. **Murata & Masui (2014, 2015b)** modified the objective function and a transition method in their simulated annealing method. Although their method (Ikeda, 2010; Murata, 2014, 2015b) can synthesize a population that has the same statistical characteristics with the real statistics, their method tried to synthesize a reduced population with only 500 or 1,000 households. The synthesized population is too small for social simulations in a real city, town or village.

In order to cope with the problems arisen in the reduced number of populations, we tried to synthesize exactly the same number of individuals in a target area such as states, counties, and prefectures using statistics of prefectures **(Murata, 2016, 2017a)**. We first increased the number of real statistics for each family type and modified a transition method (Age-Changing method) in their SA method by considering role in a family (2016). We then proposed another transition method (Age-Swap method) in their SA method that keeps the distribution of the initial population that is fit to the real statistics (2017a). Age-changing method has a better performance in reducing the error when the number of transitions in an SA method is relatively small. On the other hand, Age-swap method can reduce better than Age-changing method when the number of transitions in an SA method is relatively large.

When we increase other attributes such as geographical characteristics **(Harada, 2017)** or occupation and income **(Murata, 2017b)** to the synthesized population, populations by local governments such as city, town or village are required **(Murata, 2018)**. There are finer statistics that are statistics for each "basic unit block." The number of "basic unit blocks" in Japan is around 1.9 millions. A population synthesis method using statistics of "basic unit block" is proposed by Harada & Murata (2018). We have conducted population synthesis using the above algorithms with high performance computers in Osaka University. **Fig. 2** shows an example of a household projection on buildings in a map of Japan. Each figure in a circle shows the number of households residing in the corresponding building.

Table 1 Distributed Synthetic Populations.

Organization	Synthesized Area	Statistics
RTI International, USA	All states, USA Population: 300 million	2010 US Decennial Census 2007-2011 American Community Survey
CDRC: Consumer Data Research Center, UK	England & Wales, UK Population: (53 + 3) million	2011 UK Census
Kansai University, Japan	Japan Population: 120 million	2000 National Census 2005 National Census 2010 National Census 2015 National Census

3. Synthesized Population Distribution

Using the above synthetic methods, we have generated synthesized populations for whole Japan. We are trying to prepare the database of synthesized populations using database. There are only two organizations that distributes synthetic populations in the national level in the world. **Table 1** shows the distributed synthesized populations. Those organizations distribute nation-wide populations of their country.

Although they are distributing only one set of synthesized populations, we have synthesized 10 sets of populations now. Since any methods synthesizes populations based on the limited number of statistics, there is no guarantee that the synthesized population is exactly the same as the real population. Therefore, RSSS should be conducted on several sets of populations and find common outcome from the simulations, or a unique outcome among them. When we find a common result, it seems to be obtained from any populations with the same statistical characteristics of the real population. When we find some unique result, we should carefully see how the obtained result is caused. In order to conduct such multiple simulations, we distribute several sets of populations.

We are distributing synthetic populations with the following notations.

1) The synthetic populations do not contain any data of the real households and individuals.
2) The synthetic populations contain only the same statistical characteristics of the real households and individuals.
3) The synthetic populations do not contain any statistical characteristics that are not used in the synthetic process.
4) The synthetic population will be updated when latest statistics become available.
5) Simulations or analysis using the synthetic populations should be conducted on multiple sets of populations.
6) Outcomes of simulations and analysis should NOT be released any personal or private information that is relating to real households or individuals.

Although synthesized populations are not real populations, residents may consider that their privacy is offended by releasing

their personal information such as their occupations, income or educational back grounds. Therefore, we require researchers to conduct their simulations or analysis using multiple sets of synthesized populations in Item 5). We also require researchers not to release outcomes of their simulations or analysis in any forms that enables others to identify or estimate a private information in a certain household.

4. Further Challenges for Data Protection

In order to protect the personal or private information in the synthesized populations, we are planning to employ a cloud service that enables simulations using the synthesized populations. By employing a cloud-style service, we do not have to distribute the synthesized data themselves to researchers but allow them to access the synthesized data in their RSSSs. In order to realize such an interface for accessing the synthesized populations, we should develop online programming tools for utilizing the synthetic populations in simulations or analysis.

Another way to protect personal information is to employ secure computation (Chida, 2014). The secure computation enables users to utilize sensitive data without allowing them to see exact values of them.

5. Conclusion

In this paper, we show the current status of population synthesis of whole populations in Japan. We have developed multiple sets of synthesized populations with the same statistical characteristics of the real populations in Japan. In synthesizing the populations, we utilized the statistics conducted in 2000, 2005, 2010, and 2015. After synthesizing the populations with sex, age, kinship in their household, we are increasing attributes of each individual such as geospatial data, occupation, and income. We hope enriching such synthesized population will help researchers who try to develop real-scale social simulations or analyze micro data to see characteristics of our communities or environments.

Acknowledgement

Part of this research is funded by Foundation for the Fusion of Science and Technology in 2017, JSPS KAKENHI 17K03669 in 2017, and Tateishi Science and Technology Foundation in 2018. Synthesized populations are generated using the large-scale computing systems (VCC) of Cybermedia Center, Osaka University.

References

[Chida, 2014] K. Chida, G. Morohashi, H. Fuji, F. Magata, A. Fujimura, K. Hamada, D. Ikarashi, R. Yamamoto, Implementation and evaluation of an efficient secure computation system using 'R' for healthcare statistics, J. of the American Medical Informatics Association, Vol. 21, Is. e2, pp. 326-331, 2014.

[Davis, 1987] L. Davies: Genetic algorithms and simulated annealing; Research Notes in Artificial Intelligence, Los Altos, CA: Morgan Kaufumann, 1987.

[Harada, 2017] T. Harada, T. Murata, Projecting household of synthetic population on buildings using fundamental geospatial data, SICE Journal of Control, Measurement, and System Integration, Vol. 10, No. 6, pp. 505-512, 2017.

[Harada, 2018] T. Harada, T. Murata, Geospatial data additional method using basic unit blocks, Proc. of SICE Symposium on Systems and Information 2018, 6 pages, 2018 (in Japanese).

[Ikeda, 2010] K. Ikeda, H. Kita, M. Susukita, Estimation method of individual data or regional demographic simulations, Proc. of SICE 43rd Technical Com. on System Engineer, pp. 11-14, 2010 (in Japanese).

[Lenormand, 2013] M. Lenormand, G. Deffuant, Generating a synthetic population of individuals in households: Sample free vs sample-based methods, Journal of Artificial Societies and Social Simulation, vol. 16, no. 4, pp. 1-9, 2013.

[Murata, 2013] T. Murata, K. Konishi, Making a Practical Policy Proposal for Polling Place Assignment Using Voting Simulation Tool, SICE Journal of Control, Measurement, and System Integration, Vol. 6, No. 2, pp. 124-130, 2013.

[Murata, 2014] T. Murata and D. Masui, "Estimating agents' attributes using simulated annealing from statistics to realize social awareness", Proc. of 2014 IEEE Int'l Conf. on System, Man & Cybernetics, pp. 717-722, 2014.

[Murata, 2015a] T. Murata, N. Du, Comparing income replacement rate by prefecture in Japanese pension system, Advances in Social Simulation, pp. 95-108, 2015.

[Murata, 2015b] T. Murata and D. Masui, "A two-fold simulated annealing to reconstruct household composition from statistics", Proc. of 2015 IEEE Int'l Conf. on System, Man & Cybernetics, pp. 1133-1138, 2015.

[Murata, 2016] T. Murata, T. Harada, D. Masui, Modified SA-based household reconstruction from statistics for agent-based social simulations", Proc. of 2016 IEEE Int'l Conf. on Systems, Man, & Cybernetics, pp. 3600-3605, 2016.

[Murata, 2017a] T. Murata, T. Harada, D. Masui, Comparing transition procedures in modified simulated-annealing-based synthetic reconstruction method without samples, SICE Journal of Control, Measurement, and System Integration, vol. 10, no. 6, pp. 513-519, 2017.

[Murata, 2017b] T. Murata, S. Sugiura, T. Harada, Income allocation to each worker in synthetic populations using basic survey on wage structure, Proc. of 2017 IEEE Symposium Series on Computational Intelligence, pp. 471-476, 2017.

[Murata, 2018] T. Murata, T. Harada, Synthetic method for population of a prefecture using statistics of local governments, Proc. of 2018 IEEE Int'l Conf. on Systems, Man, & Cybernetics, pp. 1171-1176, 2018.

[Schelling, 1971] T. Schelling, "Dynamic models of segregation, Journal of Mathematical Sociology, Vol. 1, pp. 143-186, 1971.

[Wilson, 1976] A. G. Wilson, C. E. Pownall, A new representation of the urban system for modeling and for the study of micro-level interdependence, Area, vol.8, no. 4, pp. 246-254, 1976.

International Session | International Session | [ES] E-3 Agents

[3H3-E-3] Agents: safe and cooperative society

Chair: Ahmed Moustafa (Nagoya Institute of Technology), Reviewer: Takayuki Ito (Nagoya Institute of Technology)

Thu. Jun 6, 2019 1:50 PM - 3:30 PM Room H (303+304 Small meeting rooms)

[3H3-E-3-01] An Autonomous Cooperative Randomization Approach to Prevent Attacks Based on Traffic Trends in the Communication Destination Anonymization Problem

○Keita Sugiyama[1], Naoki Fukuta[1] （1. Shizuoka University）

1:50 PM - 2:10 PM

[3H3-E-3-02] Cooperation Model for Improving Scalability of the Multi-Blockchains System

○Keyang Liu[1], Yukio Ohsawa[1], Teruaki Hayashi [1] （1. University of Tokyo, Graduate school of engineer）

2:10 PM - 2:30 PM

[3H3-E-3-03] Effect of Visible Meta-Rewards on Consumer Generated Media

○Fujio Toriumi[1], Hitoshi Yamamoto[2], Isamu Okada[3] （1. The University of Tokyo, 2. Rissho University, 3. Soka University）

2:30 PM - 2:50 PM

[3H3-E-3-04] Toward machine learning-based facilitation for online discussion in crowd-scale deliberation

○Chunsheng Yang[1], Takayuki Ito[2], Wen GU[2] （1. National Research Council Canada, 2. Nagoya Institute of Technology）

2:50 PM - 3:10 PM

[3H3-E-3-05] An automated privacy information detection approach for protecting individual online social network users

○Weihua Li[1], Jiaqi Wu[1], Quan Bai[2] （1. Auckland University of Technology, 2. University of Tasmania）

3:10 PM - 3:30 PM

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

An Autonomous Cooperative Randomization Approach to Prevent Attacks Based on Traffic Trends in the Communication Destination Anonymization Problem

Keita Sugiyama*1 Naoki Fukuta*2

*1 Faculty of Informatics, Shizuoka University
*2 College of Informatics, Academic Institute, Shizuoka University

The communication destination anonymization problem is one of the problems to be resolved under some trade-offs in the cyber security field. Several approaches have been proposed for the communication destination anonymization problem such as Wang's U-TRI. However, due to the trade-offs that the user cannot take too expensive costs to make the network performance improved while keeping its security level, there remains the issues to make anonymization even over a short period of time while giving a good throughput. In this paper, we present an overview of the approach to solve this issue by introducing autonomously coordinating multiple end-hosts and a simulation environment to analyze it.

1. Introduction

When defending facilities with the camera network or patrolling ponds for avoiding illegal disposals by drones, the networks constituting them also need to be protected at the same time in order to operate them properly. It is mentioned that the anonymity of communication destination in such network is often implemented for this purpose [Wang 17]. U-TRI [Wang 17] has been proposed by Wang et al as one of the approaches for that purpose. However, it is mentioned that U-TRI still suffers from an issue when attackers are allowed to utilize their observed traffic trends [Wang 17]. In this paper, we present an overview of the approach to solve this issue by introducing autonomously coordinating multiple end-hosts and a simulation environment to analyze it.

2. Background and Related Work

2.1 Communication Destination Anonymization

It is one of the security problems on enterprise local networks that attackers are able to gather intelligence such as which end-hosts are online and which end-hosts are important by sniffing traffic in the networks. In order to prevent this problem, identifiers appear in network traffic need to be anonymized. PHEAR [Skowyra 16] and U-TRI [Wang 17] are methods to anonymize addresses in the local network. U-TRI implement anonymity by updating identifiers representing the communication destination and source in VIRO, which is a method of efficiently routing packets using the Software Defined Network, at random intervals based on idea of Moving Target Defense [Jajodia 11].

Contact: Keita Sugiyama, Faculty of Informatics, Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu-shi, Shizuoka 432-8011 Japan, cs15050@s.inf.shizuoka.ac.jp

2.2 Attacks Based on Traffic Trends

As mentioned in the original Wang's U-TRI paper [Wang 17], U-TRI leaves the problem to allow attackers to attack based on traffic trends. Although the detail is not clearly mentioned there, the following cases can happen. For example, on the system where multiple clients are managed by a server, it is expected that packets whose destination address or source address is the address of the server appear frequently since multiple clients communicate with the server. In such a case, even if the address of each end-host is updated in a certain period, it is not difficult for the attacker to identify the address of the server by investigating the appearance situation of the address in a shorter period than the address-update interval. The primary factor of that is that, U-TRI implements anonymity in the medium to long term, but anonymity is not implemented in the short term. It is possible to make that hard by making the address-update interval very short for the purpose of implementing anonymity. However, shortening the address-update interval disorderly is not a practical solution since it is expected that the network performance will be greatly impaired by increasing the packet loss rate.

In this way, U-TRI leaves the possibility of traffic analysis utilizing the fact that it is difficult to implement short-term communication destination anonymity and that the tendency of traffic tends to be biased due to the nature of the system. In this paper, we will proceed with the necessary discussion to propose a method to effectively implement short-term anonymity.

3. Overview of Proposed Approach

The aim of our proposed approach is to implement a short-term anonymity in consideration of the trade-offs with the network performance while implementing the communication destination anonymity in the medium to long term like the U-TRI does. In addition, it is also required to

	Address	The First Observed Time	The Last Observed Time
1	~~f0:00:38:4e:8f:29~~	00:00:12	00:08:12
2	~~5e:11:59:50:df:30~~	00:00:12	00:04:12
3	~~ef:fe:27:b7:3d:10~~	00:04:12	00:08:12
4	~~a0:50:88:8a:5f:33~~	00:08:12	00:12:12
5	~~bf:2c:f8:9d:db:48~~	00:08:12	00:13:30
6	**be:9a:70:5b:9f:54**	00:12:12	00:13:30

Figure 1: The recently updated addresses estimated by the attacker.

change the strategy automatically and autonomously in consideration of the current traffic trends since network traffic changes over time.

Regarding the former requirement, the approach that each end-host determines the address-update frequency according to its own importance level is considered as one of the ways of satisfying the requirements.

Regarding the latter requirement, the approach that each end-host determines the address-update frequency according to its own packet transmission/reception status and packet loss is considered as one of the methods satisfying the requirements.

Therefore, it is possible for each end-host to determine its own address-update frequency in consideration of its own importance level, packet transmission/reception status, and its potential or current level of packet losses. Here, an issue is found using this approach. The issue of this approach is that the attackers are able to predict the most recently updated address, that is, the address likely to be the address pointing to the end-host whose address is being updated frequently, by excluding addresses that have not been observed for a long time and addresses that have been observed for a long time among the observed addresses. Figure 2 shows how an attacker predicts the most recently updated address from the observed addresses. Entries 1 to 4 are the addresses that have not been observed for a long period of time. Entry 5 is an address that is being observed for a long period of time. The attacker predicts that entry 6 is the address that was most recently updated.

In this way, it becomes an issue when attackers are able to gain much profit by attacking the end-host with high frequency of address-updates if the address-update frequency can be predicted by attackers. In order to solve this issue, we also require the ability that allows each end-host to cooperate with other end-hosts for giving attackers uncertainty about their own importance level.

4. Calculation of Attack-Success Rate by Simulator

In this work, we are preparing a prototype simulator to evaluate effectiveness our approach. The prototype of simulator has a mechanism to analyze the differences among the original U-TRI and our approach regarding their abilities to prevent an attack which utilizes its poor implementation of anonymity of communication destination (Attacker A) and an attack which predicts an end-host whose update frequency of the addresses is high (Attacker B).

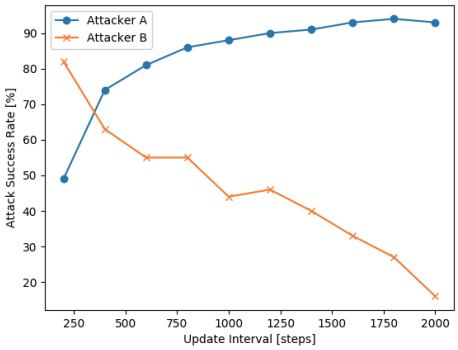


Figure 2: Attack-success rate of each attacker against the address-update interval of the server. The number of end-hosts except for the server = 6, the address-update interval of end-host except for the server = 2000 [steps].

On a situation with a network which has one server and six cameras where each camera communicates with the server, we analyze the attack-success rate of each attacker in the case where the attack-success condition is to attack the server. Figure 2 shows the results on that condition. It shows the attack-success rate when only the address-update interval of the server is changed while the address-update interval of end-hosts are fixed except for the server.

5. Conclusion

In this paper, we presented an overview of the approach to prevent attacks based on the traffic trends which is unavoidable in the original U-TRI, which provides the communication destination anonymization problem in the cyber security field by autonomously coordinating multiple end-hosts. In addition, we presented a prototype simulator to analyze differences among the original U-TRI and our approach.

References

[Jain 11] Jain, S., Chen, Y., and Zhang, Z.-L.: VIRO: A scalable, robust and namespace independent virtual Id routing for future networks, *2011 Proceedings IEEE INFOCOM*, pp. 2381–2389 (2011)

[Jajodia 11] Jajodia, S., Ghosh, A. K., Swarup, V., Wang, C., and Wang, X. S.: *Moving Target Defense: Creating Asymmetric Uncertainty for Cyber Threats*, Springer Publishing Company, Incorporated (2011)

[Skowyra 16] Skowyra, R., Bauer, K., Dedhia, V., and Okhravi, H.: Have No PHEAR: Networks Without Identifiers, in *Proceedings of the 2016 ACM Workshop on Moving Target Defense*, pp. 3–14 (2016)

[Wang 17] Wang, Y., Chen, Q., Yi, J., and Guo, J.: U-TRI: Unlinkability Through Random Identifier for SDN Network, in *Proceedings of the 2017 Workshop on Moving Target Defense*, pp. 3–15 (2017)

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Cooperation Model for Improving Scalability of the Multi-Blockchains System

Liu Keyang Ohsawa Yukio Teruaki Hayashi

Department of System Innovation, Graduate School of Engineering, The University of Tokyo

Scalability is an open question of the blockchain. Ongoing solutions, like Sharding and Side-chain, try to solve it within an independent blockchain system. We propose a cooperation model by constructing a system of multiple blockchains. In this model, secure cross chain operations can help to handle more requests. The gossip channel can help to refresh the states of other blockchains. Through manage interactions among blockchain systems, this model can limit their misbehaviors and improve scalability.

1. INTRODUCTION

Blockchain, a solution to decentralized systems, can solve problems in fields like finance[Eyal 2017], supply chain[Abeyratne 2016], crowdsourcing [Li 2018]. Currently, blockchain can provide two functions. First, a blockchain can work as a securely distributed ledger[Ren 2018]. Second, a blockchain can provide a reliable distributed calculating platform. By using smart contracts[Underwood 2016], all participants can execute functions correctly and give the same output.

Generally, a decentralized system is more robust and trusted than a centralized system. However, the scalability is its weakness. Scalability problem is the long latencies or superfluous messages caused by growing participants. Usually, it is a result of the consensus algorithm[Karame 2016]. Many solutions try to solve this problem within a blockchain system. Sidechain shifts some assets into a sidechain to realize faster responses[Back 2014]. Sharding technology tries to split participants into several shardings for parallel processing[Luu 2016]. All these works sacrifice security or consistency for the efficient responses.

This work tries to solve the scalability problem through cooperative problem-solving. In this model, each blockchain system is an independent agent. The contributions of this work are 1. A protocol for delivery versus payment(DVP) problem. 2. The framework of Blockchain's cooperation model.

2. RELATED WORKS

Sharding is an exciting idea that split blockchain into several shardings so they can handle requests simultaneously. This idea shares some similarity with multiple blockchains system. Elastico[Luu 2016] and Rapid chain[Zamani 2018] are some great implementations of this solution. In these systems, the randomness of each sharding limits the possibility of collusion and planned attacks. However, they give up security or consistency to some degree. To solve this problem, we propose a framework to weaken secure assumptions of each blockchain. By considering possible attack happens, this work focus on limiting the effect of attacks. Hence, our model allows more sacrificing of security on individual blockchain while providing better services.

Chen et al[Chen 2017] and Kan et al[Kan 2018] had finished some works about the communication between different blockchains. They simulate Internet stack and TCP protocol to create the Inter blockchain communication protocol. Although these methods are functional, they are also fragile to malicious attacks. Besides, they ignored the achievements of the consensus algorithm which is very useful in DVP problem. This work will consider the case that some blockchains are controlled or created by attackers. We can prove that these attacks cannot affect other blockchains.

3. PROBLEM MODEL

This part will clarify assumptions and notations of this model. First, $N = 1, 2, ...n$ represents the set of agents. Each agent i is a distributed network that maintains one blockchain $B_i[]$ with all terminated blocks list linearly. The participants of each agent run the consensus algorithm to maintain the blockchain and provide their services to users. Terminated block means at least $f_i > 0.5$ participants have confirmed and stored the block. f_i is the parameter of each agent's consensus algorithm.

Second, all block contents two parts: header and body. A header contains at least the hash of the previous Block, metadata of the body, and signatures of the creator. For convenience, all blockchains' contents, like transactions or Key-Value pair, is unified under an abstracted class – log. Each agent should support two operations: $verify$ and $check_i$. $verify(log, h)$ will return the validity of one log before the h^{th} block $B_i[h]$ according to the rule of the blockchain. Taking a Bitcoin's transaction as an example, input should be a subset of unspent transaction output (UTXO), and the sum of inputs should be larger than the sum of outputs. $check_i(log)$ returns the position of one log in B_i. It will return -1 when it does not exist. Hence, the

Contact: Liu Keyang, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 Japan, Department of Systems Innovation, School of Engineering,The University of Tokyo, Bldg.No.8. 507 , 070-4336-1780, stephenkobylky@gmail.com

following property held:

- a.(Validity) $\forall h > 0$ and $log \in B_i[h]$, $verify_i(log, T) = True$ for $T < h$ and $verify_i(log, T) = False$ for $T > h$.
- b.(Agreement) If $check_i(log) == True$, log can be accessed from at least f_i part of participants in agent i.

Third, we assume adding one legal log into a blockchain consume resources for all agents. Under this condition, cross chain operation becomes a DVP problem. Off-line cost's DVP problem is the job of exchanges. This work focus on the online DVP problem between two agents. Cooperation model will guarantee the payment's validity stick to the delivery of goods.

Last, we allow an agent itself can work improperly or attack some other agents. Since an agent can work independently, agents need some mechanism to control the effect of attacks. One naive way is creating a higher layer blockchain among different agents so it can be byzantine fault tolerate. This work uses another lightweight method. We create an externality of each cross-chain operation without affecting other agents. However, these externalities can prove the existence of misbehavior and punish the agent by detaching it from the network.

4. COOPERATION MODEL

This section includes the detail of the cooperation model. To begin with, we defined two extended functions for each participant of agents in our model.

4.1 Extended operation

Define a condition log $clog1 = log1||log2||j||h1||h2$ represents $log1$ is a cross chain operation related to $log2$ in agent j. The expiration height for a condition log is $h1$ and $h2$ in agent i and j. All participants of agents maintain a waiting list(WL) for cross chain log and condition log. WL supports a function $condcheck()$. When $clog1$ and $log1$ are stored together in WL, $condcheck(log1) = clog$ and $condcheck(clog1) = log1$. In other cases, it returns the existence of input log in WL.

Next, we need to define the extended function $checkEX_i(log1)$. Let $checkEX_i(log1)$=-1 if $condcheck(log1) = True$. If $condcheck(log1) = clog1$, $checkEX_i(log1) = check_i(clog1)$. In other cases,$checkEX_i(log1) = check_i(log1)$. $checkEX_i(clog1) = check_i(clog1)$

Then, we define function $verifyEx(log, h)$ in algorithm 1.

4.2 Workflow

By using previous notations, the operations to WL are following:

- When the new terminated block contains a condition log $clog1$, all participants add $clog1$ and $log1$ into their WL.

Algorithm 1 verifyEx

Input: log1 or clog1, h
Output: True or False

```
if Input is normal log then
  if condcheck(log1)! = clog1 then
    return verify(log1, h)
  else
    return checkEX_i(clog1) ≥ 0 & checkEX_i(log_1) < 0 & checkEX_j(clog2) ≥ 0 & verify(log1, h)
  end if
else if condcheck(clog)! = False then
  return False
else
  return h < h_1 & verify(log1, h)
end if
```

- Expire: When $B_i[h1]$ is terminated $clog1$ is removed from WL. When$checkEX_j(log2) \geq 0$ and $checkEX_i(log1) \geq 0$ $log1$ and $clog1$ is removed.

the workflow of cross chain operations are following:

- 1. Register: A user submits $clog1$ to one participant. Participants check $verifyEX_i(clog1, h_t)$ where h_t is the current height of blockchain B_i. If it returns True, commit $clog1$ to next block.
- 2. Condition-commit: If the newest terminated Blocks contain $clog1$, all participants add $clog1$ and $log1$ to their WL.
- 3. Pre-commit: User submits $log1$ to one participant.The participant checks $verifyEX_i(log1, h)$. If it is true, commit $log1$ to next block.
- 4. Commit: When the height of blockchain reaches $h1$, participant check $checkEX_i(log1)$ and $checkEX_j(log2)$ to determine whether to expire $clog1$.

The workflow of a success cross chain operation looks like Figure 1.

Till now, we have clarified main steps of a cross chain operation. The final step is to broadcast the latest view, like the hash of last terminated block header, of both agents. One agent can use a gossip channel to notify other agents of updating. This gossip is not necessary to be received or confirmed. However, an agent can reveal a fork by identifying an unmatched view of one agent.

4.3 Communication rule

In the cooperation model, each blockchain is an agent to act. Hence, verifying the status of one agent demands sufficient supports from its participants. Due to the property of agreement, secure connection with one agent i anchors to the parameter f_i. For a given possibility p , the required confirmations m_i should satisfied $(1 - f_i)^{m_i} < p$. Hence, $m_i \geq \frac{p}{\ln(1-f_i)}$.

A secure communication requires enough participants of agent i asks for m_j confirmations from agent j independently. The communication cost is $O(m_i * m_j)$ per time.

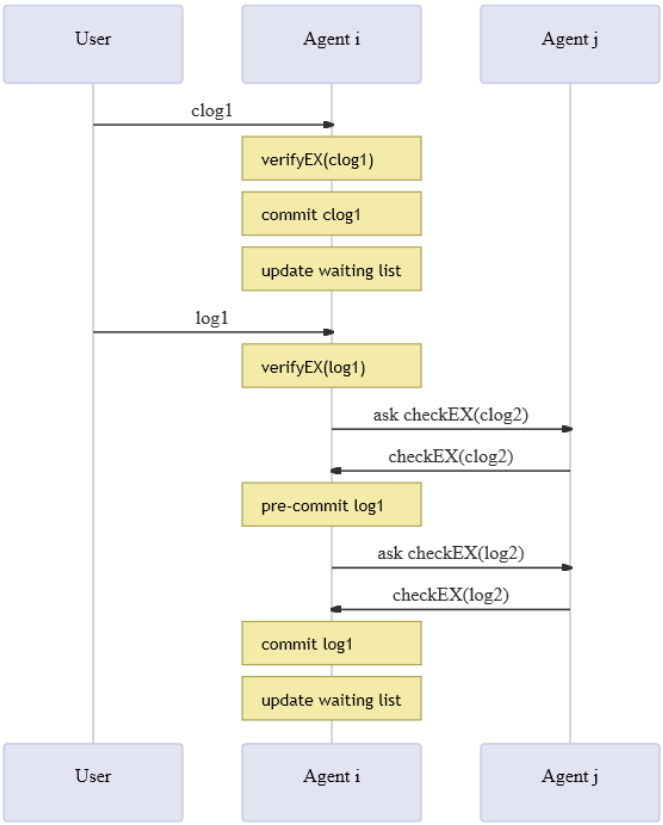


Figure 1: The work flow of one cross chain operation.

An efficient way is to select some proxies to do this communicate. Once selected proxies confirmed $checkEX_j()$, they can spread the gossip information among agent i. Under this method, the cost is $O(m_i)$. For preventing collusion, participants should randomly select proxies.

5. ANALYSIS AND EXPERIMENT

5.1 Function analysis

To view other possible situations of this protocol, we can consider the states of WL. WL can allow three statues:1. Null,2. $clog1$ and $log1$, 3. $log1$. Table 1 shows the result of VerifyEX and CheckEX, where the condition is the 3rd line in Algorithm 1. When a user submits $clog1$, only case 1 and case 3 can continue. Case 3 means new condition log is a supplement to an expired one which jumps to the final stage. When a user submits $log1$ at case 3, participants can detect that it is a rejected log. As a result, on one can admit a existed cross chain operation $log1$ without satisfying a condition $clog1$. This is the cost of agreement in a blockchain.

Assumes log' is conflicted with $log1$, which indicates the terminated blockchain can only contain one of them. The condition for committing log', $checkEX_i(log1)$ should be less than 0. Since termination requires admissions of at least half of all participants, they can exclude the possibility of conflicts within the blockchain. Once a terminated block contains log', it is impossible to activate $log1$ in case 3 any more. The reason is $clog1$ is also conflicted with log'.

Func \ WL	Null	Clog&log	log
VerifyEX(clog)	Verify(Log)	False	Verify(Log)
CheckEX(clog)	check(clog)	check(clog)	check(clog)
VerifyEX(log)	Verify(Log)	condition	Verify(Log)
CheckEX(log)	check(log)	check(clog)	-1

Table 1: Result of VerifyEX(ignore h) and CheckEX

Hence, by using the property of Blockchain, this protocol can solve the DVP problem of cross chain operations.

5.2 Security analysis

This part provides some brief proofs of security. Generally, there are two types of attacks: agent's misbehaviors and communication attacks. During the protocol, the only information needed about other agent is function $checkEX$ which affect by WL and terminated blocks. Hence, Adjusting can detect the counterfeit WL and B_i. Once attacker Eve controls the agent j, j can reply to other agents arbitrary and support any cross-chain operations. After completing a cross chain operation, Eve can create a fork to repeal the existed log. In this case, agents that received the previous view of j will anchor to the elder branch and reject new branch inherit the identity of j. The network will wait for j recovering the former branch and continue its services. Here, the gossip channel creates an externality of an agent so that it cannot change its termination within the network. The higher rate to verify gossip, the higher termination the model can propose.

Besides, there are some network attacks like Sybil attacks and DDOS attack among agents. Sybil attack means the attacker create several agents in the model to arrange attacks. However, these bot agents need to spend enough resources to convince users of other agents for one round attack. Hence, this attack is not profitable if users can manage their risk. Another way is isolating one agent like Man in the Middle(MITM) attack to block gossips. This attack is very costly when the target is a distributed network. A fixed and reliable channels can also resolve this attack. In a word, the resilience against manufactured identities depends on the value of each agent. Lastly, DDOS attacks also worth considering. Attackers can attack waiting list by creating many useless conditional logs. The solution can be charging an additional fee for registering cross chain operation. Indeed, cross chain operations require extra payment for stronger termination and complex procedures. The most significant problem relates to the gossip channel where redundancy informs can block useful gossip. Hence, the gossip channel needs some rule for filtering. Agents can require a signature for each message, limit frequency of source agents, create some periodical routes.

5.3 Experiment

This work intended to improve the scalability of one agent through cross chain operation. The experiment evaluated the average latency and gossip burden for different rates of cross chain operations in the worst case. Latency is evaluated by the number of blocks for solving same amount

of requests. Gossip represents the number of message sent in gossip channel when discard rate is 0.5. The result(Figure 2) shows a linear growth of latency with inter log rate and stable average gossip message related to the number of agents. The increasing of latency relates to the size of condition logs. In the experiment, we assume condition log spending double space compare to other logs.

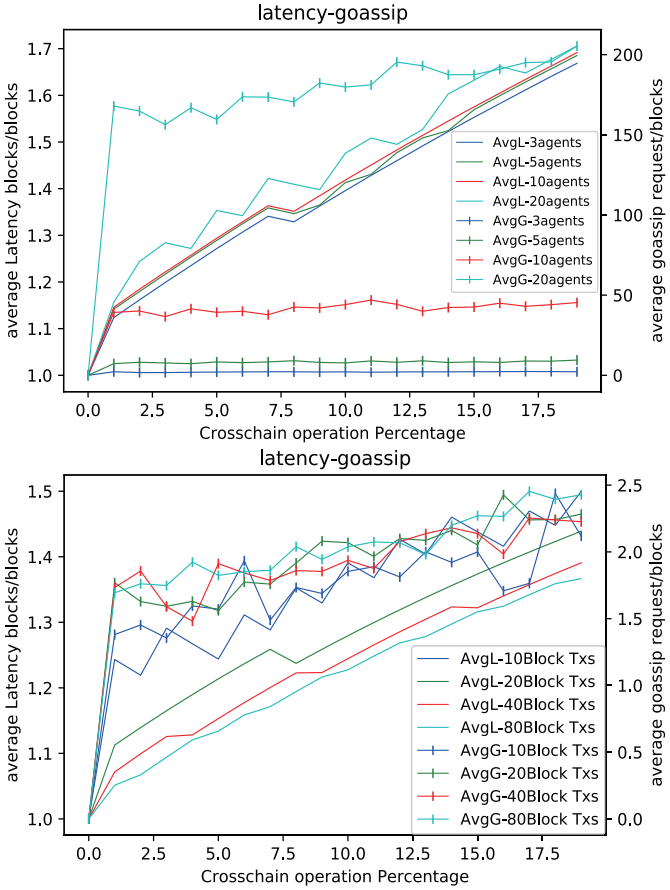


Figure 2: Average latency and gossip according to cross operation rate when transactions terminate immediately. The first picture shows the number of gossips is square to the system's scale. The second picture shows the system do not accumulate gossips or latency as time goes by.

6. CONCLUSION

This work proposes a cooperation model for improving scalability. Under the cooperation model, cross chain operation can create externality of each agent and spread it through gossip channel. The termination of one agent can partly rely on other agents and agent can pay more attention to achieve agreements. On the other hand, cross chain operation extends the ability of one agent and allow a more flexible exchange between different digital assets. We can still optimize this work in many ways. Condition log and gossip message can be compressed to reduce latency and burden of gossip channel. A gossip checking protocol can detect forks faster. The future works will focus on the consensus design under the cooperation model and optimization of gossip channel and related protocols.

7. ACKNOWLEDGMENT

This work was funded by JSPS KAKENHI, JP16H01836, JP16K12428, and industrial collaborators.

References

[Eyal 2017] Eyal I. Blockchain technology: Transforming libertarian cryptocurrency dreams to finance and banking realities[J]. Computer, 2017, 50(9): 38-49.

[Abeyratne 2016] Abeyratne S A, Monfared R P. Blockchain ready manufacturing supply chain using distributed ledger[J]. 2016.

[Li 2018] Li M, Weng J, Yang A, et al. Crowdbc: A blockchain-based decentralized framework for crowdsourcing[J]. IEEE Transactions on Parallel and Distributed Systems, 2018.

[Karame 2016] Karame G. On the security and scalability of bitcoin's blockchain[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2016: 1861-1862.

[Back 2014] Back A, Corallo M, Dashjr L, et al. Enabling blockchain innovations with pegged sidechains[J]. URL: http://www. opensciencereview. com/papers/123/enablingblockchain-innovations-with-pegged-sidechains, 2014.

[Luu 2016] Luu L, Narayanan V, Zheng C, et al. A secure sharding protocol for open blockchains[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2016: 17-30

[Ren 2018] Ren Z, Cong K, Aerts T, et al. A scale-out blockchain for value transfer with spontaneous sharding[C]//2018 Crypto Valley Conference on Blockchain Technology (CVCBT). IEEE, 2018: 1-10.

[Chen 2017] CHEN Z, Zhuo Y U, DUAN Z, et al. Inter-Blockchain Communication[J]. DEStech Transactions on Computer Science and Engineering, 2017 (cst).

[Kan 2018] Kan L, Wei Y, Muhammad A H, et al. A Multiple Blockchains Architecture on Inter-Blockchain Communication[C]//2018 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C). IEEE, 2018: 139-145.

[Zamani 2018] Zamani M, Movahedi M, Raykova M. RapidChain: scaling blockchain via full sharding[C]//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2018: 931-948.

[Underwood 2016] Underwood S. Blockchain beyond bitcoin[J]. Communications of the ACM, 2016, 59(11): 15-17.

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Effect of Visible Meta-Rewards on Consumer Generated Media

Fujio Toriumi*1Hitoshi Yamamoto*2Isamu Okada*3

*1The University of Tokyo *2Rissho University*3Soka University

Consumer Generated Media(CGM) are useful for sharing information, but information does not come without cost. Incentives to discourage free riding (receiving information, but not providing it) are therefore offered to CGM users. The public goods game framework is a strong tool for analyzing and understanding CGM and users' information behaviors. Although it is well known that rewards are needed for maintaining cooperation in CGM, the existing models hypothesize the linkage hypothesis which is unnatural. In this study, we update the meta-reward model to identify a realistic situation through which to achieve a cooperation on CGM. Our model reveals that restricted public goods games cannot provide cooperative regimes when players are myopic and never have any strategies on their actions. Cooperative regimes emerge if players that provide first-order rewards know whether cooperative players will give second-order rewards to the first-order rewarders. In the context of CGM, active posting of articles occurs if potential commenters/responders can ascertain that the user posting the article will respond to their comments.

1. Introduction

Consumer generated media (CGM) are the most active information sharing platforms in which users generate contents by voluntary participation. They include information sharing sites such as Wikipedia and TripAdviser, and question/answer forums such as Yahoo Answers. CGM reflect positive traits of the Internet because, in CGM, aggregating users' voluntary participation bears values, and thus they have network externality in which the more active users are, the more the values of the CGM are.

CGM rely on user-provided information and thus fail if information is not provided. Getting users to provide information generally requires effort costs including time costs and click costs[Nakamura 14]. Therefore, CGM users are given incentives to discourage free riding, a situation in which users receive information, but do not provide it. While huge CGM never worry about freeriding, many managers of small-sized CGM pay attention to it. CGM can be regarded as a kind of public goods game–a social dilemma game in which users may refrain from paying costs (that is, free riding), although they could benefit substantially if they contributed.

To avoid the free-rider problem, many CGM adopt incentive systems for users to receive comments as appreciation for posting articles. These comments are considered rewards for contributing to the public goods game. Moreover, many real CGM systems provide Like buttons to react to comments, which can be regarded as meta-rewards. This is because comments also give psychological benefits to original article providers as well as Like buttons give psychological benefits to their receivers.

Toriumi et al. [Toriumi 16] used a public goods game model to show that meta rewards are required to maintain cooperation. A meta reward is a reward for those who gave a reward to cooperative users. Many CGMs implement a function that allows other users to express their gratitude to those who provided information, and the users who expressed their gratitude can also be given something as a reward.

However, the model has an unrealistic hypothesis which called Linkage hypothesis: Whoever performs the first-order sanction (rewards) also performs the second-order one. This hypothesis is needed for the theoretical rationale of meta sanctions because, if the second-order sanctions are independent of the first-order sanctions, third-order free riders who shirk the second-order sanctions only are possible, and thus cooperation through meta sanctions collapses. Experimental studies have no consensus on this linkage hypothesis. Some experiments support the linkage between the first-order sanctions and cooperative behaviors [Horne 01, Horne 07] while others deny it [Yamagishi 12, Egloff 13]. The linkage hypothesis between the first-order and second-order sanctions is partially supported by an experiment of a one-shot public goods game [Kiyonari 08].

In this paper, we will model our CGM public goods game without assuming the linkage hypothesis between the first- and second-order rewards. While a previous model [Toriumi 16] uses the same parameter, r_i, as the probabilities of giving rewards and giving meta rewards, our model separates the former probability from the latter.

2. Models and Methods

In this section, we develop a model that reflects real CGM by extending the CGM model proposed by Toriumi et al. [Toriumi 12]. We then define an adaptive process of players in the model to explore feasible solutions of strategies for promoting and maintaining cooperation. Third, we introduce several scenarios to provide insight for managing real CGM by comparing their performances. Finally, we set parameter values to perform our simulation.

2.1 A restricted meta reward game model

We consider N agents playing a restricted meta reward game. The game is run for a discrete time and each period is referred to as a round. In each round, all agents play three

連絡先: Fujio Toriumi, tori@sys.t.u-tokyo.ac.jp

sequential steps in serial order. Using the case of Agent i as an example, Agent i has its own strategy denoted by (b_i, r_i, rr_i), which we will explain later.

In the first step, the agent provides its own token into a public pool with probability b_i and otherwise does not. In CGM, a contribution and a non-contribution are, respectively, regarded as an information-providing behavior and a non-providing behavior. If a token is provided by Agent i, i must pay a cost κ_0, also the other $N-1$ players receive a benefit, ρ_0.

In the second step, rewards for providing a public good may occur. In CGM, posting a comment to an information provider is regarded as a reward. If and only if Agent i provides a token, the other $N-1$ agents consider whether or not they will give a reward to Agent i. Agent $j(\neq i)$ gives a reward to Agent i with probability $p_{r_{i\to j}}$ and otherwise does not. This probability is calculated as $p_{r_{i\to j}} = \varepsilon \cdot r_j$, where r_j is j's own reward parameter and ε is an expected rate of meta rewards newly introduced in this model to consider the third challenge of the above-mentioned prior studies. If a reward is given, Agent i gains a constant benefit, ρ_1, while Agent j must pay a constant cost, κ_1.

In the third step, meta rewards for giving rewards may occur. In our model, meta rewards from contributors are possible in the first step only to consider the second challenge of the previous studies, thus making this model a restricted game. In CGM, a reply to comments is regarded as a meta reward. If and only if Agent i received a reward from Agent j, Agent i can decide whether to give a meta reward to Agent j with probability rr_i, and otherwise not. While Toriumi et.al.[Toriumi 12] assumes that $r_i = rr_i$, our model assumes that these are independent of each other to consider the linkage hypothesis. If a meta reward is given, Agent j gains a constant benefit, ρ_2, while Agent i must pay a constant cost, κ_2.

Each agent plays the above three steps four times in each round. When all agents complete these steps, each agent's final payoff at each round is regarded as its fitness value.

2.2 Simulation scenarios

In the restricted meta reward game, there is no incentive to give meta rewards, and thus players never provide meta incentives. To consider this point, we introduce players' expectations of meta rewards. We then explore how these expectations are reflected in the probability of providing rewards using the following three scenarios that are different values of expected rates of meta rewards, ε.

1. No reference ($\varepsilon = 1.0$): players do not use any reference
2. Social reference ($\varepsilon = \frac{1}{N}\sum_k rr_k$): players use the average rate of meta rewards in the group
3. Individual reference ($\varepsilon = rr_i$): players use cooperator i's probability of meta rewards

Scenario 1 is a baseline. Scenario 2 describes a situation that players can get information on a providing rate of meta rewards in CGM. For instance, we suppose a system in which seeing all meta rewards for rewards by others is possible. Scenario 3 describes a situation that visualizes a providing rate of meta rewards for information provided in CGM. In this scenario, we assume that players can decide whether or not to provide meta rewards to a cooperator after they check the providing rate of meta rewards of the focal cooperator.

Table 1: Simulation Parameters

Param	Value
N	100
Simulation steps	1000
μ (benefit-cost ratio)	2.0
δ (discount ratio)	0.8
ρ_0 (benefit of cooperation)	2.0
κ_0 (cost of cooperation)	1.0

2.3 Parameter setting

For simplicity, we set the values of the parameters above by installing two new intervening parameters: δ and μ.

$$\kappa_0 = 1.0 \tag{1}$$

$$\rho_n = \mu \cdot \kappa_n \tag{2}$$

$$\kappa_n = \delta \cdot \kappa_{n-1}, \tag{3}$$

where $n = 1, 2$.

At first, we simulate the case of $\mu = 2$ and $\delta = 0.8$ to clarify the performances of each scenario. Then, we investigate the influences of the cost-reward ratios in Section 3.2. Table 1 shows the values of the other parameters in the simulation.

3. Simulation Results

3.1 Comparison of three scenarios

We simulate 100 runs with different random seeds in each scenario, and show the averages and the variances of values using error-bars in Figs.1, 2, and 3. In these figures, the vertical axes show the step numbers while the horizontal axes show the average parameter values: Cooperation indicates cooperation rates, b_i, Reward indicates reward rates, r_i, and MetaReward indicates meta reward rates, rr_i.

As shown in Fig.1, the cooperation rate in Scenario 1 decreases at about 100 steps while increasing at the beginning. This is due to the decrease in reward rates. The rate gradually decreases immediately after the beginning and reaches 0.1 at 20 steps. No reward never bears cooperation.

Scenario 2 faces the same mechanism and thus neither scenario can maintain a cooperative regime.

In Scenario 3, on the other hand, the cooperation rate increases from the beginning, then the meta reward rate also increases and, finally, the reward rate increases, therefore maintaining a stable cooperative regime as shown in Fig.3.

Why does Scenario 3 promote cooperative regimes while Scenario 1 does not? This is quite surprising because parameter value ε is 1 in Scenario 3 while it is less than 1 in

Scenario 1. We then analyzed the time series of cooperation rates, reward rates, and meta reward rates in Scenario 3 in comparison with Scenario 1. At the beginning of the simulation, cooperative rates increased in both scenarios. However, the next phenomena are different. In Scenario 3, the meta reward rates increased before the reward rates increased. This is because players with high meta reward rates tend to receive more rewards than those with low meta reward rates. If the number of players who give rewards is sufficiently large, the high meta reward rates bear the benefit of the rewards and are larger than the costs of meta rewards. Therefore, players with high meta reward rates benefit more than those with low meta reward rates.

The more players with high meta reward rates there are, the greater the probability of receiving meta rewards when giving rewards. Therefore, players who tend to give rewards gain more benefit than those who do not, and thus the reward rates increase. High reward rates enhance the benefit of cooperation and, therefore, cooperative players have an advantage over defective players. Cooperative regimes stay robust.

3.2 Influence of cost-reward ratios

In our model, the rate of the reward benefit on the reward cost is important for promoting cooperative regimes[Toriumi 16]. Therefore, we simulated many cases with different values of μ and δ. Figure 4 shows the average rate of cooperation in 1000th step with in 50 runs per each case. In this figure, the x axis indicates μ, the y axis indicates δ, and the color bar indicates the average cooperation rates.

The scopes of μ and δ are, respectively, $0.0 \leq \mu \leq 5.0$ and $0 \leq \delta \leq 1.0$. This figure shows that

1. Cooperative regimes emerge only in Scenario 3
2. Cooperative regimes never emerge if $\mu < 1.4$ and
3. Cooperative regimes emerge if approximately $\mu \cdot \delta > 1.0$

Among these, Result 2 is consistent with a previous study [Toriumi 12] that demonstrated that cooperative regimes require a substantially large benefit of rewards compared with their costs. Our result adds the insight that it also requires a sufficiently larger value of μ in our model than the previous study's model. This is because the expected values of meta rewards are small if μ is small, and thus the incentive to give rewards vanishes.

Next, we consider Result 3. As a result of our simulation, condition $\mu \cdot \delta > 1.0$ is necessary for promoting cooperation. In terms of the relationship between rewards and meta rewards, if the benefit of meta rewards is greater than the cost of rewards, players may receive a benefit through giving rewards, and thus there are incentives to give rewards. This indicates that

$$\rho_2 > \kappa_1 \tag{4}$$

is required. If $\kappa_1 > 0$ is satisfied, equations $\rho_2 = \mu \cdot \kappa_2 = \mu \cdot \delta\kappa_1$ are satisfied, and thus the necessary condition of reward behaviors is

$$\mu \cdot \delta = \frac{\rho_2}{\kappa_1} \quad > \quad 1.0. \tag{5}$$

Strictly on this point, players do not always receive meta rewards and thus we should consider the average rate of meta rewards, $\overline{rr_i}$. Therefore,

$$\overline{rr_i} \cdot \mu \cdot \delta > 1.0 \tag{6}$$

is the necessary condition.

If this condition is satisfied, players who give rewards to other players at sufficiently large rates of meta rewards have an advantage. This also means that cooperative agents are given incentives from which they should receive a large amount of meta reward rates. This mechanism works and therefore players with large amounts of both reward rates and meta reward rates have survival advantages and, finally, cooperative regimes emerge.

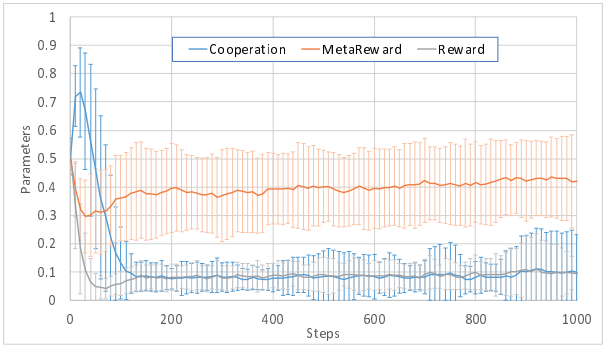


Figure 1: Result of Scenario 1

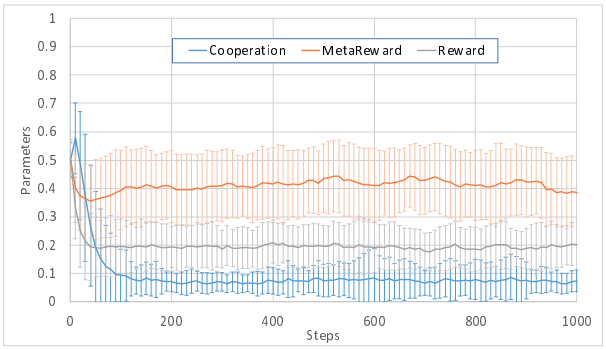


Figure 2: Result of Scenario 2

4. Discussion

While our main results support the importance of meta-rewards for activating CGM, we must state the

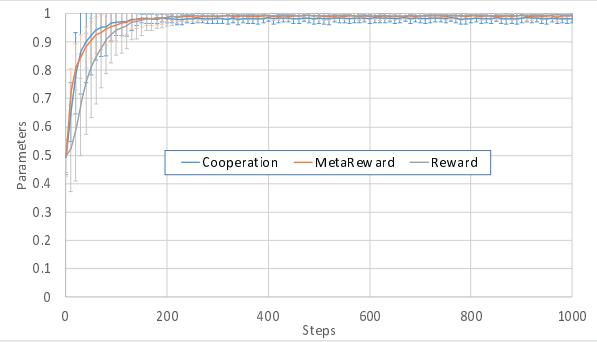


Figure 3: Result of Scenario 3

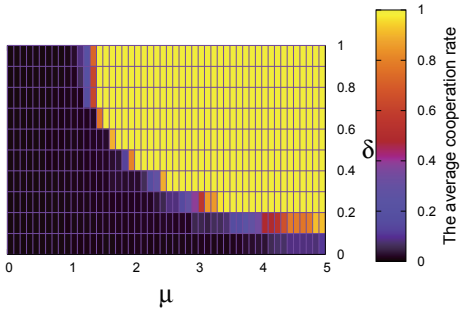


Figure 4: Change μ, δ in Scenario 3

other important drivers of real posting including brand image[Kim 16], attention seeking, communication, archiving, and entertainment[Sung 16]. Moreover, we have no option but to accept the future study on the empirical data that support that the original article providers respond to other commenters replies to sustain posting on CGM.

We developed a restricted public goods games model to overcome the mismatches found between previous models and actual CGM. Our model reveals that restricted public goods games cannot provide cooperative regimes when players are myopic and never have any strategies on their actions. Cooperative regimes emerge if players that give first-order rewards are given information that reveals whether cooperative players will give second-order rewards to the first-order rewarders. In the context of CGM, if users who post articles reply to commenters/responders, active posting of articles occurs if potential commenters/responders can ascertain that the user posting the article will respond to their comments.

This study should be extended. First, the present version of our model describes two types of players actions: cooperation as posting information and defect as non-posting. However, defect behaviors in CGM can be divided into two types: do nothing and post inadequate information. This issue should be introduced in a future version. Second, while our model assumes that all players can observe all information, this is not realistic. We are interested in the influence when the frequency of information accessibility depends on the quality of the information.

References

[Egloff 13] Egloff, B., Richter, D., and Schmukle, S. C.: Need for conclusive evidence that positive and negative reciprocity are unrelated, *Proceedings of the National Academy of Sciences*, Vol. 110, No. 9, pp. E786–E786 (2013)

[Horne 01] Horne, C.: The enforcement of norms: Group cohesion and meta-norms, *Social psychology quarterly*, pp. 253–266 (2001)

[Horne 07] Horne, C.: Explaining norm enforcement, *Rationality and Society*, Vol. 19, No. 2, pp. 139–170 (2007)

[Kim 16] Kim, A. J. and Johnson, K. K.: Power of consumers using social media: Examining the influences of brand-related user-generated content on Facebook, *Computers in Human Behavior*, Vol. 58, pp. 98–108 (2016)

[Kiyonari 08] Kiyonari, T. and Barclay, P.: Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment., *Journal of personality and social psychology*, Vol. 95, No. 4, pp. 826–42 (2008)

[Nakamura 14] Nakamura, H., Gao, Y., Gao, H., Zhang, H., Kiyohiro, A., and Mine, T.: Tsunenori Mine. Toward Personalized Public Transportation Recommendation System with Adaptive User Interface, *3rd International Conference on Advanced Applied Informatics* (2014)

[Sung 16] Sung, Y., Lee, J.-A., Kim, E., and Choi, S. M.: Why we post selfies: Understanding motivations for posting pictures of oneself, *Personality and Individual Differences*, Vol. 97, pp. 260–265 (2016)

[Toriumi 12] Toriumi, F., Yamamoto, H., and Okada, I.: Why do people use Social Media? Agent-based simulation and population dynamics analysis of the evolution of cooperation in social media, in *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 02*, pp. 43–50 (2012)

[Toriumi 16] Toriumi, F., Yamamoto, H., and Okada, I.: Exploring an Effective Incentive System on a Groupware, *Journal of Artificial Societies and Social Simulation*, Vol. 19, No. 4 (2016)

[Yamagishi 12] Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., Miura, A., Inukai, K., Takagishi, H., and Simunovic, D.: Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity, *Proceedings of the National Academy of Sciences*, Vol. 109, No. 50, pp. 20364–20368 (2012)

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Toward machine learning-based facilitation for online discussion in crowd-scale deliberation

Chunsheng Yang[*1], Takayuki Ito[*2], and Wen Gu [*2]

[*1] National Research Council Canada [*2] Nagoya Institute of Technology

The objective of this paper is to develop machine learning-based facilitation agent for facilitating online discussion in collective intelligence, particularly for online discussion in deliberation. The main idea is to model facilitator's human behaviour by using machine learning technique, case-based reasoning paradigm,. After introducing the details of the proposed machine learning-based approach for facilitation of online discussion, the paper presents some preliminary results along with some outline of the on-going research tasks and future work. The results demonstrate that it is feasible and effective to develop machine learning-based agent for smoothing the discussion and achieving a consensus.

1. Introduction

Deliberation is defined as the activity of small group of people who make the best solution for themselves [Ito 2017]. Over centuries, such decision-making process never changed. This deliberation process is controlled by a small group of powerful people who make the policies without incorporation of public opinion from crowd, and excludes the most people's involvement during the decision-making. Such an approach is becoming inadequate because many important ideas are not properly incorporated. Today, democratically, most people or crowds have to be involved in deliberation.

With the rapid development of Internet, the Internet-based online discussion in crowd-scale deliberation [Klein 2011] or in collective intelligence has attracted many efforts from researchers in social science and computer science. Online crowd decision-making support has received an amount of research interests, and some such support systems have been developed. For instance, Climate CoLab at MIT [Introne 2011] was a pioneer project which aims at harnessing the collective intelligence of thousands of people around the world to make arguments on global climate issues. The project developed a web-based crowdsourcing platform to facilitate the online argumentation [Klein 2011, Gurkan 2010, Klein 2007] democratically. Another example is COLLAGREE developed at Nagoya Institute of Technology (NiTech); it is a web-based online discussion platform [Ito, 2014], which provides a facilitator the support for managing online discussion to effectively achieve the consent through various mechanisms, including facilitation, incentives [Ito 2015], discussion-tree [Sengoku 2016], and understanding. The project team has applied the COLLAGREE to political applications such as city planning forum to collect the crowd opinion from public. For example, NiTech and Nagoya City used COLLAGREE for generating the consent for Next Generation Total City Planning. With the help of COLLAGREE, the Nagoya City gathered many opinions from public citizens. On the other hand, the people from city can understand the importance of next generation city plan. Eventually a consent decision can be achieved democratically. Such online argumentation platforms or forums require the facilitators having systematic methodologies to efficiently guild the discussion toward to consensuses by integrating ideas and opinions and avoiding flaming.

Existing online discussion systems or collective intelligence support systems require the human facilitators to conduct facilitation in order to guide/ensure the online discussion towards consensus. However, human facilitators-based online discussion systems remain several challenging issues such as human bias, time/location restriction, and human resources constrains. To address these challenges, relieve some burden of facilitators, and reuse the prior experience and skills of the facilitators, it is desirable that more advanced techniques are available for supporting the automated facilitation to achieve the consensuses efficiently.

Fortunately, the advancement of machine learning and multi-agent systems techniques provides a venue for developing facilitator agent to automate facilitations for large-scale online discussion. One of machine learning techniques available is case-based reasoning (CBR), which provides an effective reasoning paradigm for modeling the human cognition behaviors in solving real-world problems. CBR-based approach has been widely applied to many applications such as fault diagnostics [Yang 2003], recommendation systems, and judge supporting systems [Lopes 2010]. We believe in that machine learning-based facilitation, specifically, CBR-based method should be a good solution to crowd-scale deliberation or online discussion facilitation. Therefore, we propose a CBR approach to facilitating the crowd-scale online discussion in order to achieve a consensus efficiently. The main idea is to develop CBR-based facilitation actions/mechanisms, including better idea generation, smooth discussion, avoiding negative behavior and flaming, and maintaining online discussion, consensus-oriented guidance and navigation, and so on. The paper mainly discuss the basic ideas on developing machine learning-based facilitation agent and some on-going research tasks and future work.

Following this Section, the paper presents the proposed CBR-based approach for facilitating the online discussion in details;

Chunsheng Yang, National Research Council Canada, 1200 Montreal Road, Room370a@M50, Ottawa, ON, Canada, Chunsheng.Yang@nrc.gc.ca, Tel 1-613-993-0262

Section III discusses the on-going research tasks and future work; and the final Section concludes the paper.

2. Machine learning-based facilitation agent

Machine learning techniques have been widely applied to various real-world problems and have been achieved great success in developing machine learning-based modeling technologies. Today, the machine learning-based modeling technology has become a powerful tool for building models to explain, predict, and describe system or human behaviors. The main task is to develop the data-driven models from the historic data or past experience by using machine learning algorithms. The developed models have the given ability to explain, predict, and describe the system or human behaviour. For example, in the prediction applications, the machine learning-based models can forecast the system operating status, including failures or faults. With such predictions the proactive actions can be taken to maintain the system availability. In this work, we contemplate to use a case-based reasoning approach or paradigm to model facilitators' behavior or facilitation by using their experience accumulated in past.

2.1 CBR-based modeling for cognition

CBR is rooted in the works of Roger Schank on dynamic memory and the central role that a reminding of earlier episodes (cases) and scripts (situation patterns) has in problem solving and learning [Schank 1983]. Today, Case-based reasoning is a paradigm for combining problem-solving and machine learning to solve real-world problems. It has become one of the most successful applied intelligences for modeling human cognition. The central tasks in CBR-based methods and systems [Amodt 1994] are: "to identify the current problem situation, find a past case similar to the new one, use that case to suggest a solution to the current problem, evaluate the proposed solution, and update the system by learning from this experience. How this is done, what part of the process that is focused, what type of problems that drives the methods, etc. varies considerably, however". A general CBR-based system or agent can be described by a reasoning cycle composed of the following four steps:

- RETRIEVE the most similar case from existing case bases;
- REUSE the solution in the case to solve the problem such as flaming, wrong post to the issue, distraction post;
- REVISE the proposed solution if necessary;
- RETAIN the parts of this case into a case base for future problem solving.

2.2 Case composition and definition

In general a case documents relationships between problems and its solutions. CBR solves a new problem by adapting similar solutions used for a similar problem in the past. For online discussion facilitation, a case can be defined as three components (as shown as Figure 1): online discussion case description, facilitation action, and case management. Following is the brief description for each components.

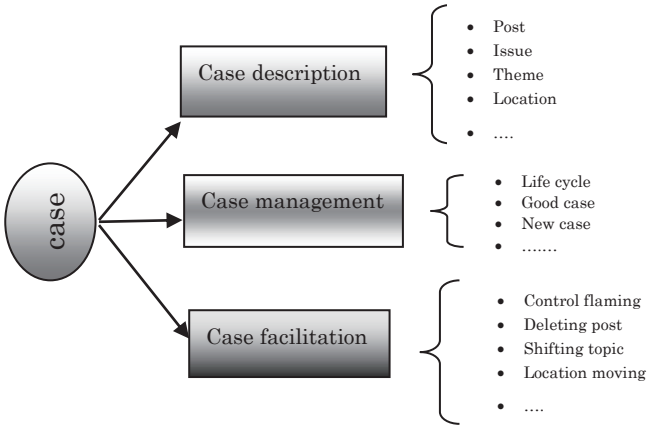


Fig.1 The case composition for online discussion

Case Description: This component contains discussion post, issue related to discussion, topics, theme, and so on. Post could be a free text, or a group of sub posts.

Case Management: This component consists of necessary case management information such as case status, case life cycle, case type, case consistence, and so on.

Case Facilitation: This component records the facilitation actions conducted by human facilitator over the past online discussion. The main facilitation could be flaming control, topic shift, post combination, post deletion, idea promotion, and so on.

From online discussion practice such as Nagoya City Planning, we have collected the data to create cases based on the case definition. It is especially useful to create facilitation cases which documents how a facilitator guided the online discussion; what kind of facilitation was used; how a facilitation action was taken, and so on.

2.3 Similarity computation

Based on the case definition above, a CBR method must provide a similarity algorithm for computing the similarity between two cases. Using computed similarity, the similar cases can be retrieved from a case base. In this work, we provide a global similarity algorithm, which computes the global similarity (sim) using Equation 1.

$$sim = \frac{\sum_{i=1}^{N} \omega_i sim_i(f_i, f_i')}{\sum_{i=1}^{N} \omega_i} \quad \cdots\cdots\cdots\cdots (1)$$

where, *sim* is the global similarity of two problems; N is the number of features or attributes that contribute to similarity; ω_i is the weigh coefficient of each feature; sim_i is a local similarity; f_i, f_i' are the i^{th} features in a case and given problem description. It is computed with Nearest Neighbor *(NN)* distance algorithm (*NN* method) for regular types of the features. For a free "text" feature, we use natural language processing techniques to compute local similarity. Particularly, we used IE (Information Extraction) method to compute the text similarity by using the library provided in OpenNPL package [Weber 2001]. We used the Maximum Entropy algorithms implemented in the OpenNLP package to compute the local similarity for two text messages as expressed as Equation 2.

$$sim_i\ (f_i, f_i') = \frac{|f_i \cap f_i'|}{(|f_i| \cup |f_i'|)} \ldots\ldots\ldots\ldots\ldots (2)$$

2.4 CBR-base facilitation agent framework

Once the case base is created, the CBR-based facilitation agent is ready constructed in online discussion system or platform. In general, such a facilitation agent can be implemented following the designed framework showed as in Table 1.

Table 1, the pseudocode for facilitation agent framework
(*Note: CB: case bases, C: case from Post, C': retrieved case from CB, FA: facilitation action from C')*

Input: *caseBase (plain text file, CB);*
Steps:
CB =Loadthecasebaseinmomorry (CB text file);
C= GetOnlineDiscussionCase (post ,issue, theme);
C'= StartReasonningCycle(C) ;
⇨ *RrtriveCase(C,SIM());*
⇨ *SolutionAdaptation(C')*
⇨ *ReviseCase(C');*
⇨ *RetainCase(C');*
FA =AdaptatFacilitation(C');
ExecuteFAcilication (FA);
Stop

As described in Table 1, the first step is to load the cases (stored in external files or database) into memory given a case composition and configuration mapping information. Once the case base is loaded into memory, a facilitation agent can execute the CBR reasoning cycles to retrieve the similar case in a case base for obtaining facilitation action give a post, issue and theme. Second step is to adapt the facilitation from the similar case for given case; the third step is to modify the case if necessary; the last step is to retain the revised case and save it back to the case base as a new case. The final step is to execute the facilitation action based on adapted facilitation action to the current post if it requires a facilitation action.

3. Discussion and Future Work

This paper mainly reports our ongoing research project. The objective is to present the ideas for developing machine learning-based agent for online discussion facilitation. Therefore many tasks are ongoing. Since focusing on CBR-based approach for facilitation, we only discuss the CBR related ongoing tasks. The other machine learning-based methods for automated facilitation will be reported in other papers.

3.1 Case structure extension and similarity algorithm

The case defined above is a basic structure. To reflect the various online discussion and complexity of facilitator's behaviour, the case structure may become complicated and complex. The similarity computation algorisms have also to be further investigated and extended from existing simple algorithm. For example, we are exploring a graph-based case structure in order to build case from a group post instead of one individual post [Gu 2018]. On the other hand, we have to investigate new algorithms for computing similarity of graph-based online discussion cases

Cases can be created either from historic data or simulation data. In this work, we conducted an online discussion forum to collect the real data. The forum was set up as a "CBR approach to support facilitation in COLLAGREE". We created the theme for an online discussion in the laboratory. The online discussion was managed and guided by a facilitator who maintains the online discussion in three phases: divergence, convergence, and evaluation. The facilitator used the support vehicle provided to navigate the forum from divergence to convergence to evaluation. Using collected data, we created some cases which reflect the facilitator's facilitation during online discussion. However, to enrich the facilitation more data are required for case creation. One way is to conduct the simulation to generate more facilitation data for creating more cases.

3.2 Machine learning-based case adaptation

In CBR research area, one remaining challenge is case adaptation. It is normal that we can't retrieve a similarity case from a case base to obtain a similar facilitation action for controlling and managing online discussion in practice. Therefore, the CBR-based agent has to adapt a facilitation action. To this end, we have to build the ability for agent to learn a new facilitation action. This motivates us to investigate the machine learning-based case adaptation methods for facilitation agent.

3.3 Case base management

This is a vital research topic for any CBR-based applications. The existing cases are manually created from the forum data collected in COLLAGREE. This is a time-consuming task and requires rich domain knowledge to understand the contents in the post or opinion. With the increasing of the collected data, manual case creation will be a challenge. An automated case creation mechanism is expected and necessary. Therefore two necessary research topics are described as follows:

(1) Automated case generation: As we discussed above, automated case creation is desirable to relieve the burden of manual case creation. From the viewpoint of machine learning, automated case creation is a supervised learning problem. It requests the annotated information to decide the case property or types. To do this sentiment analysis of the post contents is inevitable and vital for determining the case types: positive, natural, and negative. Another challenge is machine translation of language. During the online discussion it may encounter the multiple language. When generating cases from different language the automated machine translation is required.

(2) Case base management: In this work, the case base management still remains a challenge. To manage the case base efficiently, case redundancy and consistence have to be investigated in order to ensure the quality and integrity

of the case bases. Another challenge is case adaptation from the existing case and case updating to existing cases.

3.4 Validation and evaluation

Validation and evaluation for a CBR-based systems is always a challenge issue in developing CBR-based applications. It requires many efforts to design the procedures and methods. In this work, the following tasks will be conducted:

(1) Continue to collect the data from online discussion forum using COLLAGREE and create more cases for evaluation;
(2) Evaluate the performance of CBR-based systems for facilitation support by comparing the results with one from human facilities; and
(3) Validate the scalability of cases crossing different themes, even domains.

4. Conclusions

This paper reported an ongoing research project. The objective is to develop a machine learning-based facilitation agent for online discussion system to perform the automated facilitation in crowd-scale deliberation. After describing the proposed approach, we discussed some on-going research tasks and future work.

ACKNOWLEDGMENT

This work was supported by the JST CREST fund (Grant Number: JPMJCR15E1), Japan. We would like to thank all project members and related people for their contribution in the studies and the experiments. We are also grateful for all participants at Ito-Lab, Nagoya Institute of Technology in online discussion on CBR application to COLLAGREE.

References

[Amodt, 1994] A. Aamodt and E. Plaza, “Case-based reasoning: Foundational issues, methodological variations, and system approaches”, AI Communications, Vol. 7 No. i, 1994.

[Gu, 2018] W. Gu, A. Moustafa, T. Ito, M. Zhang. and C. Yang, “A Case-based Reasoning Approach for Automated Facilitation in Online Discussion Systems”, The Proceedings of The 2018 International Conference on Knowledge, Information and Creativity Support Systems (KICSS 2018), Thailand, Nov. 2018

[Gurkan , 2010] A. Gurkan, L. Iandoli, M. Klein, and G. Zollo, “Mediating debate through on-line latrge-scale atgumentation: Evidence from the field”, Information Science Vol. 180, No. 19, 2010, pp. 3686-3702

[Ito, 2015] T. Ito, Y. Imi, M. Sato, T. K. Ito, and E. Hideshima,”Incentive Mechanism for Managing Large-Scale Internet-Based Discussions on COLLAGREE”, Collective Intelligence 2015, May 31 – June 2, 2015, the Marriott Santa Clara in Santa Clara, CA (poster).

[Ito, 2014] T. Ito, Y. Imi, T. K. Ito, and E. Hideshima, “COLLAGREE: A Faciliator-mediated Large-scale Consensus Support System”, Collective Intelligence 2014, June 10-12, 2014. MIT Cambridge, USA. (poster)

[Ito, 2017] T. Ito, T. Ostuka, S. Kawasa, A. Sengoku, S. Shiramatsu, T.K. Ito, E. Hideshima, T. Matsuo, T. Oishi, and R. Fujita, “Experimental results on large-scale cyber-physical hybird discussion support”, International Jouran of Crowd Science, Vol. 1 No. 1, 2017

[Introne , 2011] J. Introne, R. Laubachar, G. Olson, and T. Malone, “The Climate Colab: Large scale model-based collaborative planning”, Proceedings of International Conference on Collaboration Technologies and Sytems (CTS 2011), 2011

[Klein, 2011] M. Klein, “Toward crowd-scale deliberation”, DOI: 10.13140/RG.2.2.12264.06401 , Massachusetts Institute of Technology, available at https://www.researchgate.net/publication/317613473Towards_Crowd-Scale_Deliberation

[Klein, 2012] M. Klein, “Enabling large-scal deliberation using attention-mediation metrics”, Computer Supported Cooperative Work(CSCW), Vol. 21, No. 4/5, pp.449-473, 2012

[Klein, 2007] M. Klein, “Achieving collective intelligence via large-scale on-line argumentation”, CCI working paper, 2007-001, April, 2007

[Lopes, 2010] E. C. Lopes and U. Schiel, “Integrating Context into a Criminal Case-based Reasoning Model”, the proceedings of 2nd International Conference on Information, Process, and Knowledge management, 2010

[Schank, 1983] R. C. Schank. “Dynamic Memory”, Cambridge Univ. Press, 1983.

[Sengoku, 2016] A. Sengoku, T. Ito, K. Takahashi, S. Shiramatsu, T.K. Ito, E. Hideshima and K. Fujita, “Discussion Tree for Managing Large-Scale Internet-based Discussion” , Collective Intelligence 2016, Stern School of Business New York University , June 1-3, 2016

[Weber, 2001] R. Weber, D. W. Aha, N. Sandhu, H. Mounoz-Avila, “A textual case-based reasoning framework for knowledge management applications”, Proceedings of Knowledge Management by Case-Based Reasioning; Experience and Management as Resue of Knowledge (CWCBR 2001), 2001

[Yang, 2003] C. Yang, R. Orchard, B. Farley, and M. Zaluski, Authoring Cases from Free-Text Maintenance Data, in Proceeding of IAPR International Conference on Machine Learning and Data Mining (MLDM 2003), Leipzig, Germany, July 5-7, 2003, pp.131-140

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

An automated privacy information detection approach for protecting individual online social network users

Weihua Li[*1], Jiaqi Wu[*1]

Quan Bai[*2]

[*1] Auckland University of Technology, New Zealand

[*2] University of Tasmania, Australia

Abstract: Massive private messages are posted by online social network users unconsciously every day, some users may face undesirable consequences. Thus, many studies have been dedicated to privacy leakage analysis. Whereas, there are very few studies detect privacy revealing for individual users. With this motivation, this paper aims to propose an automated privacy information detection approach to effectively detect and prevent privacy leakage for individual users. Based on the experimental results and case studies, the proposed model carries out a considerable performance.

1. Introduction

Online social networks (OSNs) have become ubiquitous in people's activities. The popularization of OSNs turns out to be a double-edged sword. On one hand, it provides convenience for people to communicate, collaborate, and share information. On the other hand, OSNs also come with serious privacy issues. Without given much attention by the users, a massive amount of private information can be accessed publicly through OSNs. Users may expose themselves to a wide range of "observers", which include not only relatives and close friends, but also strangers and even stalkers. This raises a serious cybersecurity issue, i.e., online privacy leak.

Online privacy leak means that an individual user shares his/her private information to people who he/she does not know well or even strangers on the Internet. This can be very dangerous for general Internet users, especially with the booming of OSNs. It is necessary to have a tool to assist general users to make better use of OSNs and protect them from leaking privacy information [Wang 11] [Hasan 13]. Hence, it is essential to detect privacy leakage in OSNs and remind individual online social network users before posting any privacy-related message. Under this motivation, in this paper, we propose a novel privacy detection framework for individual users of OSNs by using a Deep Learning approach. Twitter has been used as the source of data for training and validating our proposed framework since it is the biggest microblogging social media in the world [Mao 11]. Based on the generic definition of privacy and the characteristics of OSNs, the definition of "individual privacy" in OSNs have been formally defined. Furthermore, a deep learning-based approach has been developed and utilized to extract privacy-related entities from the messages posted by the users.

The rest of the paper is organized as follows. Section 2 reviews the existing research work regarding data leaks on OSNs. Section 3 introduces the automated privacy information detection framework. In Section 4, two experiments have been conducted to evaluate the proposed framework by using a real-world dataset collected from Twitter. Section 5 concludes this study, as well as the limitations and future work.

Contact: Dr Weihua Li, Auckland University of Technology, 55 Wellesley St E, Auckland, 1010, NZ, +64 09921 9999

2. Related Work

Privacy leakage detection in OSNs has attracted great attention to many researchers. A few studies have been conducted to analyze user privacy revealing on Twitter. People are very cautious about their personal information, e.g., home address, phone number, etc., but they consciously or unconsciously disclose their plans and activities through posting information in OSNs [Humphreys 10]. Publishing such messages online can possibly raise serious security issues. For example, a message saying "going out for holiday" implies that no one stays at home, which may cause robbery. Therefore, users should be reminded before delivering such event-related information. Mao, Shuai and Kapadia (2011) present a detection approach to analyzing three types of sensitive tweets, i.e., drunk, vacation and disease tweets. The research on privacy issues is not restricted to Twitter. Acquisti and Gross (2006) investigate the privacy concerns of users on Facebook. Dwyer, Hiltz, and Passerini (2007) compare the trust and privacy issues between Myspace and Facebook. Bhagat, Cormode, Srivastava, and Krishnamurthy (2010) show that privacy can be revealed by predicted social graph.

Whereas, very few studies investigate how to detect individual privacy information and protect individual OSNs users from online privacy leak. Therefore, in this study, instead of assisting the organizations, we target the individual online users and keep them away from privacy leakage. As almost all the posts by users are unstructured data, the information extraction plays a pivotal role in the proposed framework.

Named Entity Recognition (NER) is an important method for extracting domain-specific information [Nadeau 07]. Given the context of privacy detection domain, NER can assists users in identifying privacy-related entities after given sufficient training. Traditionally, Conditional Random Field (CRF) classifier has been employed for NER due to its robustness and reliability. Gomez-Hidalgoy et al. (2010) proposed a mechanism which is capable of detecting named entities, e.g., a company, brand, or person, using NER. Nowadays, Bi-directional Long Short-Term Memory with Conditional Random Field (Bi-LSTM CRF) model becomes more popular as it achieves more promising results [Lample 16]. Therefore, we utilize Bi-LSTM CRF for privacy-related entities extraction. Privacy Information Detection.

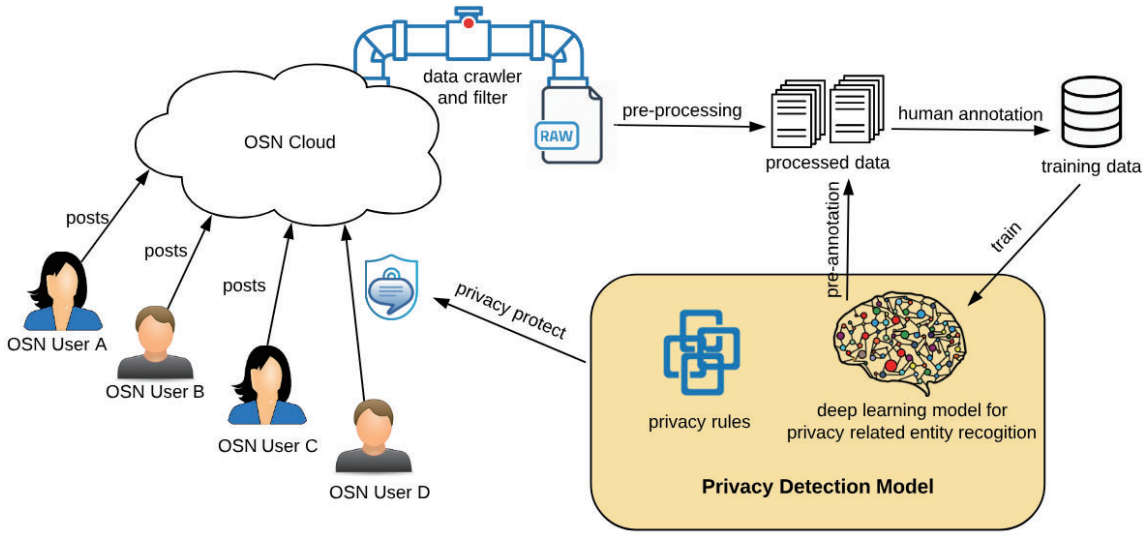


Fig 1 The Proposed Privacy Detection Framework

3. Privacy Information Detection Framework

3.1 Definition of Privacy

"Privacy" has a very broad meaning, which generally refers to the people's right to keep their personal matters and relationships secret [Gehrke 11]. In this sense, the privacy information is associated with something personal, as well as the matters of past, present, and future. Given the generic definition, in this paper, the privacy information that users tend to publish on the OSNs is defined as a sequence of words, stating or implying any individual's personal information, preferences, events that he or she involved.

Based on the definition given above, the privacy information incorporates four categories of entities, i.e., PERSON, TRAIT, PREF, and EVENT. More specifically, PERSON refers to any expression that identifies a real person; TRAIT represents the personally identifiable information, such as birth date and phone number; PREF refers to an individual's preference or hobbies; EVENT indicates the matters or activities that one involves anytime anywhere. Therefore, given a word sequence, the judgment of privacy information can be summarized as a rule as follows:

$$\exists \mathrm{PERSON} \wedge (\exists \mathrm{TRAIT} \vee \exists \mathrm{PREF} \vee \exists \mathrm{EVENT}) \longrightarrow \mathrm{P(message)}$$

3.2 The Automated Privacy Detection Framework

Our automated privacy information detection framework is demonstrated in Figure 1. The framework illustrates how the privacy detection model gets trained and utilized.

Users keep posting messages to the OSNs hosting in the cloud. Such raw unstructured and public data can be obtained through crawlers or APIs provided by the OSNs. For example, Twitter allows developers to search public tweets if the proposed project is approved. Given the context of privacy detection, the potential privacy-related data should be filtered and downloaded. Pre-processing is conducted based on specific rules, such as removing meaningless words and characters and parsing word sequences to tokens. The processed data are supposed to be further enriched by running through the pre-annotation if a privacy-detection NER model is available. Next, human involvements, i.e., manual annotation, are required. Specifically, according to the aforementioned definition of privacy information, it is essential to recognize the privacy-related entities, i.e., PERSON, TRAIT, PREF, and EVENT. The annotation also aims to figure out these four types of entities from the processed data. The annotated dataset is then fed into the deep learning model for training.

The privacy detection model consists of two components, i.e., a deep learning model for privacy-related entities recognition and privacy definition rules. For any posting messages by the OSN users, the proposed model is capable of judging whether the message is privacy-related or not. Moreover, as the privacy rules are properly defined, the privacy detection model can also explain the reason why the message is potentially privacy-related. Using a single deep learning model for private messages classification definitely loses the capability of justification.

3.3 Privacy-Related Entities Recognition

The privacy-related entities recognition plays an important role in the entire framework. There are two major aspects affecting the performance of an NER model, i.e., the annotation approach and the algorithm.

In this study, the Bi-LSTM CRF model has been employed for privacy-related entities recognition in our model, as it is capable of achieving more promising results compared with that of other classic algorithms when being applied to NER [Lample 16]. Bi-LSTM can learn long-term dependency due to the structure of the 'cell' in the hidden layer. Moreover, it can adjust the impact of previous states on the current states through the forget gate, input gate and output gate in the 'cell' [Graves 05]. However, it lacks the feature analysis on the sentence level, which can be solved by CRF. It can consider contextual conditions to make global optimal predictions. Combining the LSTM and CRF together can label sequence effectively when ensures to extract contextual features [Huang 15].

In regards to the annotation approach, BIO encoding scheme is utilized to tag entities in NER task [Kim 04]. BIO encoding scheme is a standard method which can solve the joint segmentation problem in labelling sequence by transforming them into raw labelling problem. Specifically, 'B-' is used as a prefix of an entity, implying the beginning of an entity; prefix 'I-' tags other characters indicating the tag is inside of an entity and 'O' is used for characters which do not belong to any pre-defined entities. For example, privacy-related entities fall into BIO scheme are normally annotated as follows:

I	*watch*	*a*	*movie*	*with*	*Christine.*
B-PERSON	**B-EVENT**	**I-EVENT**	**I-EVENT**	**O**	**B-PERSON**

4. Experiments

Two experiments have been conducted to evaluate the proposed privacy detection framework. The first experiment aims to train a privacy-related entities recognition model using Bi-LSTM CRF model. The second experiment gives some case studies to further demonstrate the effectiveness of the proposed privacy detection model.

4.1 Data description

Twitter is one of the largest OSNs, which enables users to conduct online social activities, including the distribution of any ideas or information. In Twitter, the messages that are posted and interacted by users are known as "tweets". Twitter provides APIs, allowing developers to search and store tweets. Therefore, we utilize Twitter API to collect 18k tweets by searching for some terms which potentially result in privacy leakage, such as pronouns, sensitive words, plans, etc.

4.2 Experiment 1

In Experiment 1, a privacy detection model based on Bi-LSTM CRF is trained to recognize the privacy-related entities. Through which, the users can be prompted before potential privacy leakage occurs. According to the definition of privacy and BIO encoding scheme mentioned previously, nine tags have been defined, i.e., 'B-PERSON', 'I-PERSON', 'B-TRAIT', 'I-TRAIT', 'B-PERF', 'I-PERF', 'B-EVENT', 'I-EVENT' and 'O'. Around 200 tweets have been annotated manually by applying these nine tags.

In this experiment, we leverage three traditional evaluation metrics as follows:

- **Precision**: the percentage that privacy-related entities can be labelled correctly among all the entities which are labelled privately in the test dataset.
- **Recall**: the percentage that privacy-related entities can be labelled correctly among all the actual privacy entities in the test dataset.
- **F1-score**: the weighted average of precision and recall, which takes both the two measures into account.

After 50 epochs' training, the performance of the deep learning model is demonstrated in Table 1.

Table 1 Performance of Privacy-Related Entities Recognition

Entity	Precision	Recall	F1-score
PREF	0.99	0.67	0.8
TRAIT	0.68	0.88	0.77
PERSON	0.98	0.93	0.95
EVENT	0.77	0.67	0.71
Avg/Total	**0.86**	**0.83**	**0.84**

4.3 Experiment 2 Case Study

In this experiment, we further demonstrate the effectiveness of the proposed privacy detection model by selecting three tweets posted recently and analyzing the results produced by the model.

Case 1: Adam and I are having lunch tomorrow.

Results: Adam (B-PERSON) and I (B-PERSON) are having (B-EVENT) lunch (I-EVENT) tomorrow.

Explanations: Based on the privacy rules, this tweet is privacy-related since it mentions both PERSON and EVENT.

Case 2: Watching a movie is a good way to relax!

Results: Watching (B-EVENT) a (I-EVENT) movie (I-EVENT) is a good way to relax!

Explanations: This tweet is just a simple statement regarding "Watching a movie", which is not a private one.

Case 3: My son is crazy about coke.

Results: My (B-PERSON) son (I-PERSON) is crazy about coke (B-PREF).

Explanations: This tweet talks about PERSON and PREF, it is privacy-related.

5. Conclusion and Future Work

In this paper, we presented a privacy information detection framework for individual OSN users. The objective is to protect end users from potential privacy leakage before posting any messages. The proposed framework explains the process of data collection, processing, model training and how it works. Both privacy rules and Bi-LSTM CRF model are leveraged in the privacy detection model. Thus, the proposed model is equipped with the capability of both detection and results explanation.

This study is still very preliminary and there is huge space for further investigation and extension. In the future, we intend to utilize a larger training dataset for performance evaluation and improve the performance of the privacy-related entities recognition by fine-tuning the parameters of Bi-LSTM CRF. Moreover, different tweets are associated with different degrees of privacy leakage. How to evaluate and score the privacy-leakage degree is also under our consideration.

References

[Wang 11] Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P. G., & Cranor, L. F. "I regretted the minute I pressed share: A qualitative study of regrets on Facebook.". *Proceedings of the seventh symposium on usable privacy and security*,pp.10 (2011).

[Humphreys 10] Humphreys, L., Gill, P., & Krishnamurthy, B. "How much is too much? Privacy issues on Twitter." *Conference of International Communication Association* (2010).

[Mao 11] Mao, H., Shuai, X., & Kapadia, A. "Loose tweets: an analysis of privacy leaks on Twitter." *Proceedings of the 10th annual ACM workshop on Privacy in the electronic society*. (2011).

[Hasan 13] Hasan, O., Habegger, B., Brunie, L., Bennani, N., & Damiani, E. "A discussion of privacy challenges in user profiling with big data techniques: The excess use case." *Big Data (BigData Congress), 2013 IEEE International Congress on*. (2013).

[Wang 17] Wang, Q., Bhandal, J., Huang, S., & Luo, B. "Classification of private tweets using tweet content." *Semantic Computing (ICSC), 2017 IEEE 11th International Conference on.*(2017).

[Aborisade 18] Aborisade, O., & Anwar, M. "Classification for Authorship of Tweets by Comparing Logistic Regression and Naive Bayes Classifiers." *2018 IEEE International Conference on Information Reuse and Integration (IRI)*. (2018).

[Lample 16] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. "Neural architectures for named entity recognition." *arXiv preprint arXiv:1603.01360* (2016).

[Bengio 94] Bengio, Y., Simard, P., & Frasconi, P.. "Learning long-term dependencies with gradient descent is difficult." *IEEE transactions on neural networks* Vol.5, No.2, pp.157-166 (1994).

[Graves 05] Graves, A., & Schmidhuber, J. "Framewise phoneme classification with bidirectional LSTM and other neural network architectures." *Neural Networks* Vol.18, No.5-6, pp. 602-610 (2005).

[Huang 15] Huang, Z., Xu, W., & Yu, K. "Bidirectional LSTM-CRF models for sequence tagging." *arXiv preprint arXiv:1508.01991* (2015).

[Bhagat 10] Bhagat, S., Cormode, G., Srivastava, D., & Krishnamurthy, B. "Prediction Promotes Privacy in Dynamic Social Networks." *WOSN*. (2010)

[Acquisti 06] Acquisti, A., & Gross, R. "Imagined communities: Awareness, information sharing, and privacy on the Facebook." *International workshop on privacy enhancing technologies.* (2006)

[Dwyer 07] Dwyer, C., Hiltz, S., & Passerini, K. "Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace." *AMCIS 2007 proceedings,* pp. 339 (2007).

[Kim 04] Kim, J. D., Ohta, T., Tsuruoka, Y., Tateisi, Y., & Collier, N."Introduction to the bio-entity recognition task at JNLPBA." *Proceedings of the international joint workshop on natural language processing in biomedicine and its applications.* pp.70-75 (2004).

[Wu 15] Wu, Y., Xu, J., Jiang, M., Zhang, Y., & Xu, H. "A study of neural word embeddings for named entity recognition in clinical text." *AMIA Annual Symposium Proceedings*, Vol. 2015, p. 1326 (2015).

[Nadeau 07] Nadeau, D., & Sekine, S. "A survey of named entity recognition and classification." *Lingvisticae Investigationes*, Vol.30, No.1, pp.3-26. (2007).

[Gehrke 11] Gehrke, J., Lui, E., & Pass, R. "Towards privacy for social networks: A zero-knowledge based definition of privacy." *Theory of Cryptography Conference*, pp. 432-449 (2011).

[Gomez-Hidalgo 10] Gomez-Hidalgo, J. M., Martin-Abreu, J. M., Nieves, J., Santos, I., Brezo, F., & Bringas, P. G. (2010, August). Data leak prevention through named entity recognition. In Social Computing (SocialCom), 2010 IEEE Second International Conference on (pp. 1129-1134). IEEE.

[3J3-E-4] Robots and real worlds: Human Interactions

Chair: Yihsin Ho (Takushoku University), Eri Sato-Shimokawara (Tokyo Metropolitan University)
Thu. Jun 6, 2019 1:50 PM - 3:10 PM Room J (201B Medium meeting room)

[3J3-E-4-01] Automatic Advertisement Copy Generation System from Images

○Koichi Yamagata[1], Masato Konno[1], Maki Sakamoto[1] （1. The University of Electro-Communications）

1:50 PM - 2:10 PM

[3J3-E-4-02] Eye-gaze in Social Robot Interactions

Koki Ijuin[2], ○Kristiina Jokinen Jokinen[1], Tsuneo Kato[2], Seiichi Yamamoto[2] （1. AIRC, AIST Tokyo Waterfront, 2. Doshisha University）

2:10 PM - 2:30 PM

[3J3-E-4-03] A Team Negotiation Strategy that Considers Team Interdependencies

○Daiki Setoguchi[1], Ahmed Moustafa[1], Takayuki Ito[1] （1. Nagoya Institute of Technology）

2:30 PM - 2:50 PM

[3J3-E-4-04] Identity Verification Using Face Recognition Improved by Managing Check-in Behavior of Event Attendees

○Akitoshi Okumura[1], Susumu Handa[1], Takamichi Hoshino[1], Naoki Tokunaga[1], Masami Kanda[1] （1. NEC Solution Innovators, Ltd.）

2:50 PM - 3:10 PM

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Automatic Advertisement Copy Generation System from Images

Koichi Yamagata[*1], Masato Konno[*1], Maki Sakamoto[*1]

[*1] Graduate School of Informatics and Engineering, The University of Electro-Communications

When we want to sell something, the presence of a good advertisement copy often affects sales. In this research, we develop an automatic advertisement copy generation system. Most existing systems only enter keywords, and the potential image of the product is not necessarily reflected by keywords only. We propose a method to generate advertisement copies using images as input to convey potential messages and the world view. This method uses Word2vec and color information of ad images, and both were confirmed to be effective by evaluation experiments. In the evaluation experiments, the mean score of the proposed method was significantly larger than 4 out of 7, and most of the subjects answered positively to ad copies of our method.

1. Introduction

Advertisement copies describe the features of products with a short number of characters and impact sentences, which are factors that greatly contribute to building brands such as companies and promoting purchase willingness of products. In recent years, there have been many researches on the generation of Japanese sentences and advertisement copy generation. In existing researches, however, they did not focus on describing potential messages and the view of the world that the producer wanted to convey.

In this research, we propose a method to generate appropriate copies reflecting color and sensitivity of given images by using a database that maintains the relationship between color and words.

There are also several prior studies on the relationship between color and sensitivity. Iiba et al. [Iiba 13] focused on the relationship between color and word sensibility, and constructed a system that recommends appropriate colors and fonts for textual sensibility images considering the effect reminiscent of the color of words. Further, Nakamura et al. [Nakamura 12] focused on the relationship between color and lyrics and constructed a music retrieval system with color as input. In this research, we focus on the relationship between colors and words and aim to automatically generate advertisement copies with color input.

Recently, in the field of natural language processing, various methods handling words as distributed expressions are increasing. Mikolov et al. [Mikolov 13] developed Word2vec which is a new model that improves the precision of vector operation of words considering the similarity between words in sentences. In this research, we utilize Word2vec to extract synonyms from a large corpus.

Existing researches on advertising copying were mainly methods of inputting words and keywords, and they were not taken into consideration to convey the latent message or the world view that the creator wished to convey. On the other hand, in this research, we construct an automatic advertisement copy generation system using images as inputs. Color information is extracted from the input image, and a large number of advertisement copies are generated by using a database that maintains the relationship between words and colors, from which adequate copies are determined as outputs.

Contact: Koichi Yamagata, UEC, 1-5-1, Chofugaoka, Chofu, Tokyo182-8585, Japan, koichi.yamagata@uec.ac.jp, 042-443-5535

2. Methods

The flow of the proposed method is as follows:

1. Color information contained in given image is extracted, and some keywords are extracted from the database storing the relationship between words and colors.
2. From the lyrics database, some phrases including the keywords extracted in step 1 are searched and extracted as sentence templates.
3. By using a deep neural network (dnn) based classification model that classifies words used for ad copies or not trained by ad copies corpus, word candidates are selected from the word extracted in step 1.
4. The noun in the template sentences extracted in step 2 are replaced by the words selected in step 3, and they are taken as copy candidate sentences.
5. The ad copy candidate sentences are evaluated by the similarities given by Word2vec, and the most evaluated ad copy candidate sentence is outputted.

In step 1, we utilize the word-color database constructed by [Konno 18]. This database was constructed based on the idea of the relation between sensitivities and music/colors studied by [Nakamura 11]. In this database, among the songs released between 1968 and 2017, we defined words that were often used for lyrics as primitive words (PWs), and each PW has a 45-dimensional color information vector obtained by a psychological experiment. Figure 1 shows 45 colors to define the 45-diensional color vector space. For words other than PWs that were not psychologically tested, latent semantic analysis was performed on them, and color information vectors were given based on the similarities with PWs.

In step 3, pytorch is used as a dnn library. The input data is a bag-of-words vector of a sentence to be classified, and the output data is “copy” or “not-copy”.

In step 4, we use Word2vec model which learned 100,000 songs of Japanese lyrics corpus, and which contains similarities with a poetic point of view. Using this model, we calculate the similarities between words (nouns, verbs, adjectives etc.) contained in sentences and evaluate candidate sentaQSZQences of ad copies.

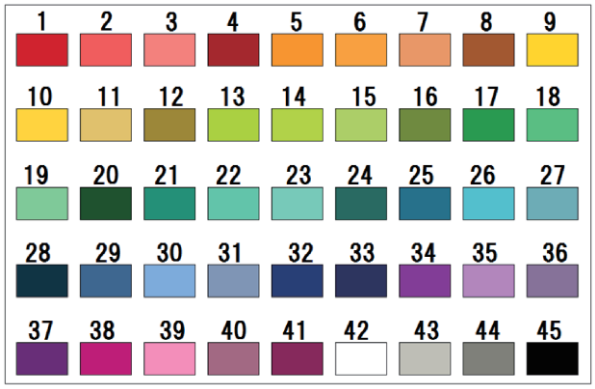


Fig. 1: 45 colors to define the 45-dimmensional color vector space.

3. System Evaluation

In order to evaluate our system, we conducted an evaluation experiment with 30 subjects (7 females and 23 men, mean age = 22.63). We adopted 20 advertisement images actually used from four categories (beverage and food, travel, beauty, fashion), and we prepared five ad copies for each advertisement image by five methods:

A) original ad copy
B) proposed method
C) proposed method without Word2vec
D) proposed method without using the image
E) RNN

For example, when input image was a picture of three can coffees against the background of the night sky, an ad copy of method B (translated into English) was “You will surely drink it tonight”, method C outputted “You laugh, even the deadly midnight dropping the mountain, forever”, method D outputted “Eat coffee and coffee”, method E outputted “Feel coffee and go”.

Subjects were asked to evaluate a total of 100 advertisement copies for the images including the original ad copies and the ad copies outputted by the proposed method. The subjects evaluated the following questionnaires on a scale from 1 to 7:

i. Is it appropriate as an ad copy in this category?
ii. Is it appropriate as an ad copy regardless of category?
iii. Is the grammar of the ad copy appropriate?
iv. Does the ad copy follow the impression of the image?

Table 1 shows the results of the questionnaires (averages and standard errors of scores). We can see that the mean score of the proposed method B is much larger than 3, and most of the subjects answered positively to our method. Comparing the methods A and B, we can see that the proposed method is slightly less than the scores of original advertisements. However, it is not so bad and the difference is small. In the food category, it is confirmed that the proposed method has smaller score differences with respect to the original advertisement. Comparing the methods B and C, we can see that the use of Word2vec is effective. Comparing the methods B and D, we can see that the use of images is effective. Comparing the methods B and E, we can see that the proposed method is much better than the RNN method.

Table 1: Results of questionnaires to evaluate each method (averages and standard errors of scores)

	i	ii	iii	iv
A	5.29±0.07	5.75±0.06	6.03±0.06	5.37±0.07
B	4.90±0.07	5.17±0.06	5.55±0.06	4.22±0.08
C	2.82±0.07	3.15±0.08	3.31±0.08	2.51±0.06
D	4.23±0.08	3.92±0.08	3.81±0.09	3.20±0.07
E	3.18±0.07	2.74±0.07	2.05±0.06	2.75±0.07

4. Conclusion

This study proposed a method to generate advertisement copies using images as input to convey potential messages and the world view. This method uses not only Word2vec but also color information of images, and both were confirmed to be effective by evaluation experiments. In the evaluation experiments, the mean score of the proposed method was significantly larger than 4 out of 7, and most of the subjects answered positively to our method. The proposed method was slightly less than the scores of original advertisements. However, it was not so bad and the difference was small.

References

[Iiba 13] Iiba, S., Doizaki, R., and Sakamoto, M.: Color and Font Recommendations based on Mental Images of Text, Transactions of the Virtual Reality Society of Japan, 18(3), pp.217-226 (2013)

[Nakamura 12] Nakamura, T., Utsumi, A., and Sakamoto, M.: Music Retrieval Based on the Relation between Color Association and Lyrics, Transactions of the Japanese Society for Artificial Intelligence, Volume 27, Issue 3, pp. 163-175, 2012.

[Mikolov 13] Mikolov, T., Yih, W., Zweig, G.: Linguistic Regularities in Continuous Space Word Representations, Conference of the North American Chapter of the Association for Computational Linguistics: Human Language, Technologies (NAACL-HLT-2013) (2013)

[Konno 18] Konno, M., Suzuki, K, and Sakamoto, M.: Sentence Generation System Using Affective Image, 2018 Joint 10th International Conference SCIS and 19th ISIS, 678-682.

[Nakamura 11] Nakamura, T., Kawanishi, K., and Sakamoto, M.: A Possibility of Music Recommendation Based on Lyrics and Color, Transactions of the Japanese Society for Artificial Intelligence, Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Volume J94-A, pp.85-94, No.2 (2011)

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Eye-gaze in Social Robot Interactions – Grounding of Information and Eye-gaze Patterns

Koki Ijuin[*1] Kristiina Jokinen[*2] Tsuneo Kato[*1] Seiichi Yamamoto[*1]

[*1] Doshisha University [*2] AIRC, AIST Tokyo Waterfront

This paper examines human-robot interactions and focuses on the use of eye-gaze patterns in evaluating the partner's understanding process. The goal of the research is to understand better how humans focus their attention when interacting with a robot and to build a model for natural gaze patters to improve the robot's engagement and interaction capabilities. The work is based on the AIST Multimodal Corpus which contains human-human and human-robot interactions on two different activities: instruction dialogues and story-telling dialogues. The preliminary experiments show that there are differences in the eye-gaze patterns given expected and non-expected responses, which affects their understanding and grounding of the presented information. The paper corroborates with the hypothesis that eye-gaze patterns can be used to predict grounding process and provide information to the speaker about how to proceed with the presentation, so as to support the partner's understanding and building of the mutual knowledge. Some consideration is given to future improvements in methodology.

1. Introduction

The goal of the research is to understand better how humans focus their attention when interacting with a robot and to build a model for natural gaze patters to improve the robot's engagement and interaction capabilities. The work follows from the pilot study (Jokinen 2018) in which human gaze patterns were studied when they interacted with a humanoid Nao robot using the WikiTalk application (Jokinen and Wilcock 2014) which allowed the user to navigate among Wikipedia topics, and is also related to eye-gaze in second-language learning with robots (Fujio et al. 2018).

In this paper, we focus on eye-tracking technology and its use in instruction giving and story-telling activities. The hypothesis examined in the paper is that there is a difference between the interlocutor's eye-gaze patterns depending on how their understanding proceeds, i.e. if the partner's utterance is understood, misunderstood or non-understood. By measuring eye-gaze activity in the communicative context we build a model that enables estimation of the partner's level of understanding, and consequently, modification of the presentation if the partner eye-gaze signals problems in the grounding of information. Such a model can help the humanoid robot to better tailor its presentations to the human user, i.e. to enable the use of the partner's eye-gaze signals to establish an appropriate way to continue. In particular, it will enable us to study how eye-gaze is used in grounding information and creating mutual understanding of the discussion topic.

Smooth interaction requires that the partners can easily understand each other and are able to build their conversation on mutual knowledge of what has been discussed. The process of creating such mutual knowledge is called grounding, i.e. the partners ground the semantics of their utterances in the context of their interaction and the context of their world knowledge, Clark & Wilkes-Gibbs (1986).

Earlier work shows that in social situations, humans are sensitive to another person's gaze: it constructs new shared knowledge, communicates experiences, and creates social linkages (Argyle & Cook, 1976; Kendon, 1967). Visual attention is important in cognitive studies (Skarrat et al. 2012), and e.g. turn-taking is commonly coordinated by gaze (Jokinen et al., 2013). Broz et al. (2015) provides an overview of the work on eye gaze and human robot interaction.

Contact: Kristiina Jokinen, AIRC AIST Tokyo Waterfront, Koto, Aomi, Tokyo, *kristiina.jokinen@aist.go.jp*

2. Data collection setup

Experiments were set-up using the lab's SMI eye-tracker and the Nao robot, to collect eye-tracking data. Fig 1. Shows the setup. Each participant had two conversations, one with a human partner (HHI) and one with a humanoid robot (HRI). The conversations were about 10 minutes long. One of the experimenters played the role of the human partner but was different from the one who gave instructions to the participant.

The experiment was conducted in Japanese or English depending on the participant's preferred language. The instructions were the same for both HRI and HHI conditions. Before the experiments, the participants signed a consent form and filled in a pre-experiment questionnaire of their background and expectations. After each interaction (HRI and HHI), they filled in another questionnaire focussing on their experience in the interaction.

Data consists of 30 participants (20 Japanese, 10 English), each having both HHI and HRI conversation. The participants (10 female) were students and researchers, age 20-60, with experience on IT, but no experience on robots. Of the participants, 14 had instruction and 16 chat dialogues.

Data analysis has started using the standard gaze frequency and duration measurements, annotations and statistical analysis,

to analyse the user's gaze patterns during interaction with the robot and with a human partner. Special attention is paid to gesturing and head nodding, and conversational instances such as turn-taking, feedback, and problem cases.

3. Annotation and Analysis

3.1 Annotation Method

Annotation for duration of utterances were done with automatic silence segmentation of ELAN. The audio files which were recorded with eye tracker were used to annotate utterances. Automatic silence segmentation was conducted two times with different thresholds of loudness for determining the silence. The one with low threshold annotates both participant's and partner's utterances, and the other one with high threshold annotates only participant's utterances. The values of those thresholds were manually set by each conversation. The segmentation of partner's utterances was calculated by subtracting those automatic segmentations.

After the segmentation, the participant's utterances in human-robot conversations were manually classified into four types from the perspective of robot's feedback to that utterances: Correct-Understood, Miss-Understood, None-Understood, and Other. Correct-Understood (CU) were tagged to the utterance which robot recognized and gave the correct feedback, Miss-Understood (MU) were tagged to the utterance which robot recognized but gave the unexpected or wrong feedback, and None-Understood (NU) was tagged to the utterance which robot did not recognized and did not give any feedback.

To annotate the eye gaze activities of participants automatically, we created the robot detection system with OpenCV3. We used cascade classifier to detect the position of robot's face in video recorded with eye sight camera of eye tracker. The robot's face was detected as rectangle, and the rectangle of robot's body was estimated with the position of robot's face. After detecting the robot's face and body, the eye gaze activities were automatically annotated into two groups; Gaze Face and Gaze Body, by judging whether the coordinates of gaze point captured by the eye tracker were in those detected rectangles or not. Fig 2 shows the result of robot detection system and gaze point of the participant.

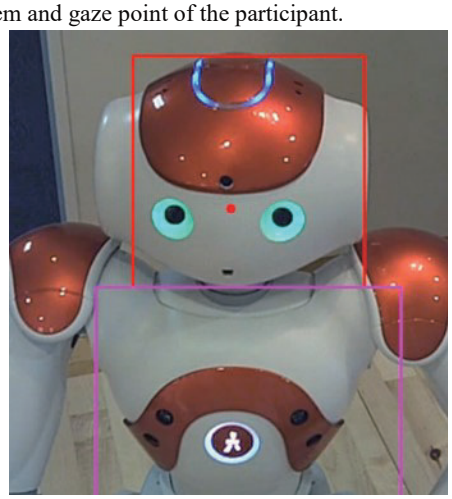


Figure 2 Snapshot of the result of robot detection system. Red rectangle represents the robot's face, purple rectangle represent, the estimated robot's body, and red dot represents the point of participant's gaze

3.2 Methodology of analyses of eye gaze activities

In order to verify how the participant uses his/her eye gaze activities, we conducted quantitative analyses of eye gaze activities during utterances, pauses just before the beginning of utterances, and pauses just after the end of utterances.

We used Gazing Ratios in order to understand how the participant uses eye gaze activities during the human-robot conversations.

Gazing Ratio is defined as:

$$\frac{1}{N}\sum^{N}\left(\frac{DG_{(i)}}{\text{duration of } i\text{-th } w}\right.$$

where $DG_{(i)}$ represents that the duration of participant's gaze toward the robot during *i*-th window. Three types of windows were used: before utterance window, during utterance window, and after utterance window. *N* represents the total number of windows. Gazing Ratio was calculated for each utterance type (CU, MU, and NU).

4. Preliminary Results

The Gazing Ratios were calculated with the data of 19 participants. Figure (ppt p. 26) shows that the temporal change of Gazing ratio for each utterance type in human-robot conversations. The results of eye gaze activities show that the participants tend to gaze away from the robot after they finishes speaking regardless of correctness of robot's feedback. After the robot gives feedback, the participants shift their gaze to the robot again. However, when the robot does not give any feedback to the participants, the participants keep gaze away to the robot for a while, and then they gaze at the robot again.

These results suggest that after the participants answer the

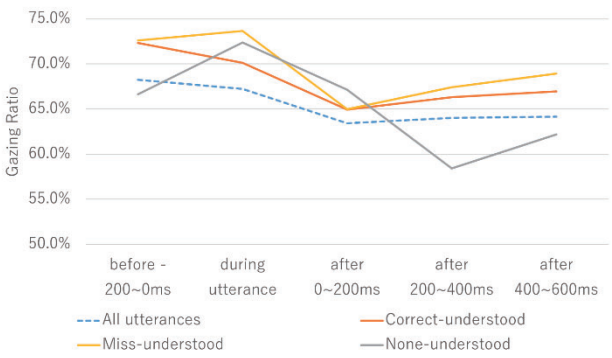


Figure 3 Gazing Ratios of participant during utterances, before the beginning of utterances, and after the end of utterances

question from the robot, they gaze away from the robot during the interval that the robot starts speaking, and if the robot does not start speaking, the participants soon realize that there is something wrong with the conversation so that they gaze at the robot in order to monitor what is going on to the robot.

This quantitative difference of eye gaze activities according to the robot's reaction might be useful to predict the participant's state whether he/she is waiting for the robot's feedback or not.

5. Conclusions

This paper has presented preliminary studies concerning eye-gaze in human-robot interaction and focused especially on the understanding of the presented information. Such grounding is important for the human-robot interaction to progress smoothly and for the robot to exhibit context-aware capability, i.e. be able to take the user's multimodal signals in the given conversation al context into account. The work is based on the AIST Multimodal Corpus which includes eye-tracking data on natural interactions between two humans and between a human and a robot.

The work continues on further analysis and annotation of the corpus and building computational models of the use of eye-gaze in signaling the interlocutors' understanding. As the corpus contains both human-human and human-robot interactions in similar conversational situations, further research is focused on studying the differences in the human gaze behaviour in the two types of conversational settings. This will deepen our knowledge of the function of gazing in interaction in general and the role of being able to detect and analyse gaze-patterns also in human-robot interactions.

Acknowledgement

This paper is based on results obtained from *Future AI and Robot Technology Research and Development Project* commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

References

Argyle, M. and Cook, M. (1976). Gaze and Mutual Gaze. Cambridge University Press, Cambridge.

Broz, F., Lehmann, H., Mutlu, B., and Nakano, Y. (Eds.) (2015). Gaze in Human-Robot Communication. John Benjamins Publishing Company.

Clark, H. & Wilkes-Gibbs D. Referring as a collaborative process. *Cognition 22:*1-39, 1986

Fujio, S., Ijuin, K., Kato, T., Yamamoto, S. (2018). Measurement of Gaze Activities of Learners with Joining-in-type RALL System, Proceedings of the 2018 IEICE General Conference, March 2018 (In Japanese)

Jokinen, K., Furukawa, H., Nishida, M., Yamamoto, S. (2013). Gaze and Turn-taking behaviour in Casual Conversational Interactions. ACM Transactions on Interactive Intelligent Systems (TiiS) Journal, Special Section on Eye-gaze and Conversational Engagement, Vol 3, Issue 2.

Jokinen, K. and Majaranta, P. (2013). Eye-Gaze and Facial Expressions as Feedback Signals in Educational Interactions. In D. Griol Barres, Z. Callejas Carrión, R. López-Cózar Delgado (Eds.) Technologies for Inclusive Education: Beyond Traditional Integration Approaches. Chapter 3, pp.38-58. Hershey, PA: Information Science Reference, IGI Global.

Jokinen, K. and Wilcock, G. "Multimodal Open-domain Conversations with the Nao Robot." In Natural Interaction with Robots, Knowbots and Smartphones - Putting Spoken Dialog Systems into Practice. Springer Science+Business Media, 2014. pp. 213-224

Kendon, A. (1967). Some functions of gaze–direction in social interaction. *Acta Psychologica*, 26 (1): 22–63.

Skarratt, P.A. et al. 2012. Visual cognition during real social interaction. *Frontiers in human neuroscience*. 6, (Jan. 2012), 196

A Team Negotiation Strategy that Considers Team Interdependencies

Daiki Setoguchi[*1] Ahmed Moustafa[*1] Takayuki Ito[*1]

[*1] Nagoya Institute of Technology

ABSTRACT

In automated negotiation, team negotiation is poised as one of the most important negotiation techniques. A team is a group where multiple interdependent agents participate in negotiations as a negotiating party during the course of negotiations. Existing team negotiation strategies did not consider the change in decisions due to the interdependencies within the team. In this paper, we propose a team negotiation strategy that considers the interdependencies within the same team. Towards this end, in the proposed negotiation strategy, we first set the parameters that represent the interdependencies that exist within the team in both directions. Thereafter, a voting process is performed in each direction. By weighting the degree of dependency of each team member, a change in agent decision due to its dependency relationships is measured. A comparative experiment with the existing team negotiation strategies showed the efficiency of the proposed strategy.

1. Introduction

Numerous researches on automated negotiation agents have been done in multiagent research fields [1, 2]. Negotiation has emerged as an important social activity wherein different people with different targets seek to reach an agreement in order to satisfy each other's interests and is indispensable in the real world where various goals exist. Therefore, automated negotiation is attracting attention as it can bring the benefits of negotiation in order to solve real life problems. In this regard, it is thought that it can be applied to route change interference and scheduling system in e-commerce system and transportation system [1, 2, 3]. One of the most important automated negotiation techniques is team negotiation [4, 5]. A team is a group where multiple interdependent agents participate in negotiations as a negotiating party during negotiations. There are numerous negotiation scenarios between multiple groups in the real world such as negotiations between a couple and a real estate agent, a negotiation between a friend and a travel agency.

In these scenarios, the team in the negotiation participates in the negotiation as a single party, but cannot be regarded as one agent. This is because they may have internal conflicting preferences when making team decisions. Even if one of the team members is unlikely to accept the proposal from the negotiating partner, this proposal may be compromised and accepted when another member accepts it. In this way, since all the team members have individual preference information and are affected by the decision of other agents, linear preference information cannot be expressed as one agent. Therefore, it is important to negotiate with the dependency amongst the team members.

However, the existing team negotiation approaches did not consider interdependencies within the team during negotiation [4, 5]. As teams are interdependent by nature in team negotiations, interdependencies must be considered during these negotiations. For example, when an agent in a team accepts the opponent's proposal, the agent that depends on that agent becomes more likely to accept the same opponent's proposal by the degree of dependency. In this context, it becomes necessary to evaluate these changes in decisions due to dependencies within the team.

In this paper, we propose a team negotiation strategy that considers the dependency relationships within this team. The proposed negotiation strategy, bidirectionally sets the parameters that express these dependency relationships that exist within the team. By setting bidirectional dependencies within the team, it becomes possible to express unilateral dependencies. In each mechanism, when the voting process is performed, by weighting the dependency degree of each team member, change of agent decision according to the dependency relation is expressed.

2. Proposed Team Negotiation Strategy

First, we define interdependencies within the team as follows. Let parameter $w(a, b)$ define Agent A 's dependency on Agent B, and each team member sets parameters to all other team members according to its dependency relationships. This is defined as the degree of dependency from agent A to agent B. When there are N team members, the dependency $W(a)$ that Agent A receives can be defined by Equation (1).

$$W(a) = \Sigma\, w(i, a) \quad (1)$$

Where $w(i, a)$ is the degree of dependency that Agent A receives from Agent i. For example, when there are three team members, their interdependencies can be represented as shown in Figure 1.

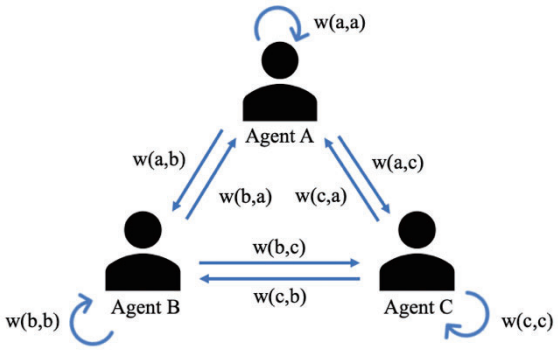


Figure 1: Dependency relationships amongst three team members

As shown in Figure 1, we can see that the dependencies between Agent A and Agent B, Agent B and Agent C, Agent A

Contact: Daiki Setoguchi, Nagoya Institute of Technology, setoguchi.daiki@itolab.nitech.ac.jp

and Agent C can be defined from both directions. By this, it can also be defined in a state that depends unilaterally.

Members with a high degree of dependency who receive it can be regarded as members with high decision and speaking rights within the team because other members are strongly influenced by their opinion. Therefore, members with high dependence are members of high importance within the team, and if the utility value of members with high importance is high, other members can easily compromise. Conversely, members with low importance depend heavily on other members, so they depend on the utility values of other members. Since the importance value in the team at this time has a difference in importance depending on the members, the weighted sum by importance rather than the sum of the utility values of simple team members is the utility value of the team. For example, utility values of members with high importance are simply added with a simple summation, but other members are influenced by the dependency within the team, and the utility value increases. Therefore, it is necessary to calculate the increment of the utility value due to the dependency relationship. Therefore, when the utility value of Agent A is assumed to be $u(a)$, the utility value of the team can be defined by Equation (2).

$$U = \Sigma\, u(i)\, W(i) \tag{2}$$

Where $u(i)$ is the utility value obtained by Agent i. $W(i)$ is the importance level of Agent i and can be regarded as the decision power that Agent i has in the team. In this research, we propose a novel team negotiation strategy that maximizes the utility value U of the whole team.

2.1 Accept/Reject Opponents' Offer

The Accept mechanism for a team determines whether to accept the proposed bid from the negotiating partner. Towards this end, all team members need to vote regarding the proposed bid by the opponent. Let $s(a)$ be the acceptance function employed by Agent A in order to assess the proposed bid. $s(a)$ is a function that returns 1 when Agent A chooses acceptance, and 0 if not. Here, the acceptance function is defined by Equation (3).

$$f = \Sigma\, s(i)\, W(i) \tag{3}$$

Where $s(i)$ represents s is a variable that returns 0 or 1 as to whether or not the Agent i accepted. $W(i)$ is the importance level of Agent i. In other words, it is synonymous with $W(i)$ in the team judged to be accepted when Agent i accepted. Therefore, when voting is done, it is necessary to obtain the importance degree of the accepted member as the voting right and the acceptance function as the sum of the importance degree. As a result of voting in this way, accept as a team if the acceptance function exceeds half the size of the team. Since the acceptance function is the sum of the importance of the received members in the team, when it exceeds half the size of the team, members who exceed the majority in the team accepted the other party's bid. Therefore, accept the proposal of the negotiating partner as a team's decision. Otherwise, the team starts the offer proposal mechanism.

2.2 Offer Proposal

The Offer mechanism decides and transmits the bid to be proposed to the negotiating partner as a team. The proposed approach employs a voting mechanism that selects widely accepted candidates such as Borda count [6]. Submit the bid that each team member wishes to propose within the team. After that, each team member evaluates all the submitted bids by using its own utility function. Let the utility value obtained when Agent A accept Agent B's proposed Bid is $u'_{A(b)}$. Agent A has a dependence on agent B for $w(a, b)$, and it is necessary to weight it when evaluating it depending on the degree of dependency. The utility value $u_{A(b)}$ for Agent B's proposed Bid for Agent A is as follows. Next, we evaluate and rank all the bids submitted within the team by Equation (4).

$$u_{A(b)} = u'_{A(b)}\,\{1.0 + w(a, b)\} \tag{4}$$

Where $u'_{A(b)}$ is the utility value when accepting B's proposed. Also, $w(i, a)$ is the degree of dependency that Agent A receives from Agent i. Assign the score from the set [0, |A| - 1] to the submitted bid along with its ranking. | A | is the total number of bids submitted. All team members make this ranking, and the highest score bid is sent to the negotiating partner as the team's proposed bid.

3. Experiment And Discussion

In this experiment, we use GENIUS which is a general-purpose negotiation platform as evaluation environment [7]. GENIUS is an open source software, aimed at negotiation simulation and the development of automated negotiation agents. Since GENIUS supports Java API that is necessary for agent development, development becomes easy with basic knowledge of Java programming. We set the experiment setting as follows. The Automated Negotiating Agents Competition (ANAC), an international competition of automated negotiation agents, employs this simulator, where several researches on automated negotiation agents are actively conducted. In addition, the negotiation problems used in the past ANAC competition have been prepared as a standard, and they support the development of an effective negotiation strategies in various negotiation problems. We set the experiment setting as follows.

- We set up 3 team members and use 3 agents of Atlas 3, Caduceus, PonPokoAgent.
- We also used Farma as a negotiating partner. All the agents used for the experiments received high ranking results at ANAC competitions.
- We used two parts, partydomain and Domain 8, implemented in GENIUS as a negotiation domain.
- Negotiating with setting the maximum time of negotiations to 180 turns, the agents negotiated 10 times for each team member's permutation. That is, 3! × 10 = 60 automated negotiations were made in one negotiation domain.
- We set the following two interdependencies in the team.

	Agent A	**Agent B**	**Agent C**
Agent A	0.5	0.2	0.3
Agent B	0.3	0.6	0.1
Agent C	0.2	0.1	0.7

Table 1: Dependency within the team

	Agent A	Agent B	Agent C
Agent A	0.6	0.2	0.2
Agent B	0.3	0.5	0.2
Agent C	0.3	0.3	0.4

Table 2: Dependency within the team

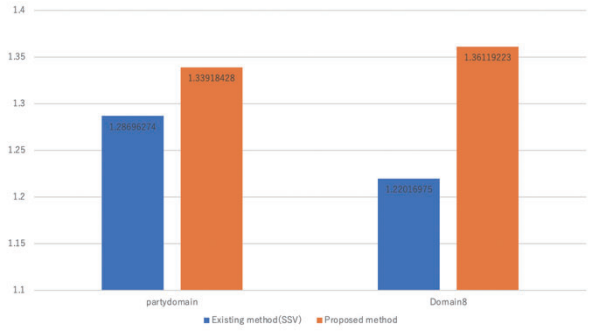


Figure 2: Average negotiation result (Table1)

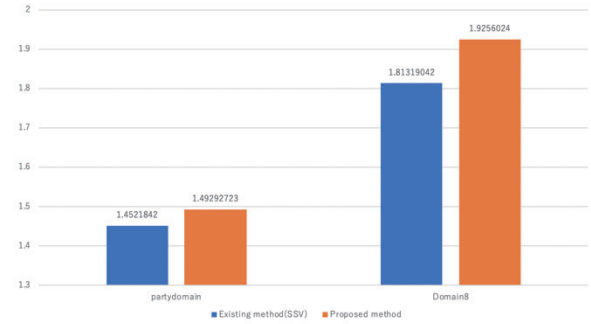


Figure 3: Average negotiation result (Table2)

The results of the experiment are shown in Figure 2 and Figure 3. As demonstrated in Figure 2, the results show that the score increases by about 0.05 for partydomain and about 0.14 for Domain 8. Since the number of team members is three, the increase in score per person is increased by about 1.8% for party domain and about 4.7% for Domain 8. As demonstrated in Figure 3, the results show that the score increases by about 0.04 for partydomain and about 0.11 for Domain 8. Since the number of team members is three, the increase in score per person is increased by about 1.3% for party domain and about 3.7% for Domain 8. Table 3 shows the results of examining the number of agreement proposal candidates in order to investigate the difference in increment by the negotiation domain.

Domain Name	Issues	Total Bid
partydomain	6	3072
Domain8	8	6561

Table 3: Comparison of size of negotiation space

From Table 3, when comparing partydomain and Domain 8, the total number of bids is larger in Domain 8. As the total number of Bids increases, the options of the proposed Bid spread, so the probability of agreeing on the same agreement decreases and the difference in the agreement proposal based on the strategy is largely reflected. From Figure 2, Figure 3 and Table 3, it can be seen that the utility value of the team is increasing as the total number of agreement proposals is larger. Therefore, as the total number of agreement proposals increases, the difference between the utility values of existing teams and the proposed method teams increases, so it can be concluded that the proposed method adapts to the utility value of the team that changes according to the dependency relationship.

4. Conclusions And Future Work

In this paper, we proposed a novel negotiation strategy that considers the interdependencies within the team in team negotiation scenarios. Towards this end, the proposed strategy implements appropriate considerations by setting the parameters that represent the interdependencies within the team, and then, using these parameters to weigh the relationship in each direction. A comparative experiment with the existing team negotiation strategies demonstrated the efficiency of the proposed strategy. As for future research, planned to investigate how to quantify the parameters that represent the interdependencies in actual negotiation problems. In addition, it become very difficult to simulate the actual negotiation scenarios unless there is a method to formulate appropriately from these scenarios. As another research direction, we also plan to study the change in the dependency score due to the change in the negotiating partner. As the negotiation partner changes, the result changes accordingly, so we plan to investigate how the agreement proposals change with the proposed approach.

References

[1] Kanamori, R., Takahashi, J. and Ito, T. “Evaluation of Traffic Management Strategies with Anticipatory Stigmergy”, Journal of Information Processing, Vol.22, No.2(2014).

[2] Sen, S. and Durfee, E. H. “On the design of an adaptive meeting scheduler.”, Artificial Intelligence for Applications, 1994., Proceedings of the Tenth Conference on. IEEE (1994).

[3] Kraus, S. “Strategic Negotiation in Multiagent Environments”, MIT press (2001).

[4] Sánchez-Anguix, Víctor, V., Botti, V., Julián, V., & García-Fornes, A. “Analyzing intra-team strategies for agent-based negotiation teams.” The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

[5] Sanchez-Anguix, V., Julian, V., Botti, V., & Garcia-Fornes, A. “Reaching unanimous agreements within agent-based negotiation teams with linear and monotonic utility functions.” IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 42.3: 778-792, 2012.

[6] Nurmi, Hannu. “Voting systems for social choice.” Handbook of Group Decision and Negotiation. Springer, Dordrecht, 2010. 167-182.

[7] Lin, R., Kraus, S., Baarslag, T., Tykhonov, D., Hindriks, K., & Jonker, C. M. “Genius: An integrated environment for supporting the design of generic automated negotiators.” Computational Intelligence 30.1 (2014): 48-70.

The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Identity Verification Using Face Recognition Improved by Managing Check-in Behavior of Event Attendees

Akitoshi Okumura[*1] Susumu Handa[*1] Takamichi Hoshino[*1] Naoki Tokunaga[*1] Masami Kanda[*1]

[*1] NEC Solution Innovators, Ltd.

This paper proposes an identity-verification system using continuous face recognition improved by managing check-in behavior of event attendees such as facial directions and eye contact (eyes are open or closed). Identity-verification systems have been required to prevent illegal resale such as ticket scalping. The problem in verifying ticket holders is how to simultaneously verify identities efficiently and prevent individuals from impersonating others at a large-scale event at which tens of thousands of people participate. We previously developed Ticket ID system for identifying the purchaser and holder of a ticket. This system carries out face recognition after attendants check-in using their membership cards. The average face-recognition accuracy was 90%, and the average time for identity verification from check-in to admission was 7 seconds per person. The system was proven effective for preventing illegal resale by verifying attendees of large concerts; it has been used at more than 100 concerts. The problem with this system is regarding face-recognition accuracy. This can be mitigated by securing clear facial photos because face recognition fails when unclear facial photos are obtained, i.e., when event attendees have their eyes closed, are not looking directly forward, or have their faces covered with hair or items such as facemasks and mufflers. In this paper, we propose a system for securing facial photos of attendees directly facing a camera by leading them to scan their check-in codes on a code-reader placed close to the camera just before executing face recognition. The system also takes two photos of attendees with this one camera after an interval of about 0.5 seconds to obtain facial photos with their eyes open. The system achieved 93% face-recognition accuracy with an average time of 2.7 seconds per person for identity verification when it was used for verifying 1,547 attendees of a concert of a popular music singer. The system made it possible to complete identity verification with higher accuracy with shorter average time than Ticket ID system.

1. Introduction

Identity verification is required in an increasing number of situations. Let us take an example of a case in which many people are admitted to an event. It used to be that in such cases, having a document, such as a ticket or an attendance certificate, checked was sufficient to gain entry; the need for personal authentication was not seriously considered due to the limited amount of time for admitting all participants. Many events with high ticket prices had designated seating, so it was not necessary to assume that some tickets may have been counterfeit. However, the advent of Internet auctions in recent years has made it easier to buy and sell tickets at the individual level. This has resulted in an increase in illegal ticket scalping, i.e., tickets being purchased for resale purposes. Equity in ticket purchasing is required not only by ticket purchasers but also by event organizers and performers [Chapple 16]. Consequently, event organizers have had to deal with complaints about malicious acts by undesignated individuals who take advantage of fans by buying and selling tickets on the Internet. In many cases, therefore, any ticket buying and selling outside the normal sales channels is prohibited. Ticket-sales terms now often stipulate that tickets are invalid when people apply for them using a pseudonym or false name and/or false address or when they have been resold on an Internet auction or through a scalper. Illegally resold tickets have in fact been invalidated at amusement parks and concert halls [JE 15]. Verification has therefore become a more important social issue than ever before. The problem in verifying ticket holders is how to simultaneously verify identities efficiently and prevent individuals from impersonating others at a large-scale event at which tens of thousands of people participate. To solve this problem, we previously developed Ticket ID system that identifies the purchaser and holder of a ticket by using face-recognition software [Okumura 17]. Since the system was proven effective for preventing illegal resale by verifying attendees at large concerts of popular music singers and groups, they have been used at more than 100 concerts. However, it is necessary to improve face-recognition accuracy because face recognition fails when unclear facial photos are obtained, i.e., when event attendees have their eyes closed, not looking directly forward, or have their faces covered with hair or items such as facemasks and mufflers. We propose an identity-verification system for attendees of large-scale events using continuous face recognition improved by check-in behavior of event attendees such as facial directions and eye contact (eyes are open or closed).

Akitoshi Okumura, NEC Solution Innovators, Ltd.
2-6-1 Kitamikata, Takatsu-ku, Kawasaki, Kanagawa 213-8511

2. Ticket ID System Using Face Recognition

2.1 Outline of Ticket ID System

Thorough verification for preventing individuals from impersonating others is in a trade-off relationship with efficient verification. The problem in verifying ticket holders is how to simultaneously verify identities efficiently and prevent individuals from impersonating others at a large-scale event in which tens of thousands of people participate. The solution should be suitable within practical operation costs for various

sized events held in various environments including open air. As a practical solution combining efficiency, scalability, and portability for a large-scale event, we developed Ticket ID system, which consists of two sub-systems, a one-stop face recognition system (one-stop system) and a check-in system [Okumura 17]. The one-stop system uses the high-speed and high-precision commercial face recognition product NeoFace [NEC 17]. The one-stop system is implemented in a commercially available tablet terminal, and the recognition result is displayed with regard to the facial-photo information of 100,000 people within about 0.5 seconds. The check-in system supports identity verification of attendees. A venue attendant checks in by placing his/her membership card on the card reader and initiates face recognition by the taking of his/her photos. The following steps make up the ticket-verification procedure from ticket application to admission [Okumura 17]:

Step 1: Tickets to popular events are often sold on a lottery basis at fan clubs or other organizations where membership is registered. Individuals applying for tickets register their membership information as well as their facial photos. In the same way for an ordinary ID photo, the registered facial photo is a clearly visible frontal photo taken against a plain background. The face must not be obstructed by a hat, sunglasses, facemask, muffler, or long hair.

Step 2: Event organizers notify ticket winners, i.e., successful applicants that have been selected.

Step 3: On the day of the event, venue attendants receive membership cards from attendees, and use a card reader to verify that attendants entering the venue are successful applicants at the event venue, as shown in Fig. 3.

Step 4: The attendants use the one-stop system to confirm that the photo taken at the time of application and the collation photo show the same person. The attendants explain the verification through face recognition to the attendees and instruct them where to stand in front of the terminal. Then, they execute the face-recognition process using the terminal to confirm the attendees are those who applied for the tickets.

Step 5: The admission procedure is carried out in accordance with the face-authentication results.

2.2 Problems with One-stop System

The average time for identity verification from check-in to entry admission was 7 seconds per person, and the average accuracy of face recognition was 90%. It is necessary to improve face-recognition accuracy by securing clear facial photos because face recognition fails when unclear facial photos are obtained, i.e., when event attendees have their eyes closed, are not looking directly forward, or have their faces covered with hair or items such as facemasks and mufflers. When face recognition fails, venue attendants have to verify attendees carefully by direct visual inspection. This increases the mental and physical burden on attendants, which makes attendees have an unreliable impression of the system. When face-recognition accuracy is 90%, two attendees are successively verified without face recognition failure with a probability of 81%. This means that 19% of attendees may experience face-recognition failure or observe it in front of them. Improving face-recognition accuracy is critical for decreasing attendants' stress and attendees' waiting time.

3. Continuous-Face-Recognition System

3.1 Managing Check-in Behavior of Attendees

We propose an identity-verification system for attendees of large-scale events using continuous face recognition improved by managing check-in behavior of the attendees. The proposed system enables attendees to check in themselves (check-in doers are not attendants, but attendees). While the previous system is equipped with a card reader, the proposed system verifies attendees with a QR code reader set up at the same position for recognizing faces of attendees standing still in front of a venue attendant, as shown in Fig. 4. Managing facial directions and eye contact are two major issues regarding facial recognition. The proposed system addresses these issues with the following methods:

1) Managing facial direction

The proposed system secures facial photos of attendees directly facing a camera by leading them to scan their QR codes just before executing face recognition. We found most people spontaneously look at the code-reader, i.e., turn their faces to the reader during check-in. The face-recognition camera of the proposed system is placed at the same position as the code-reader, as shown in Fig. 4, which makes it possible to take an attendee's photo when directly facing the camera when the photo is taken just after check-in.

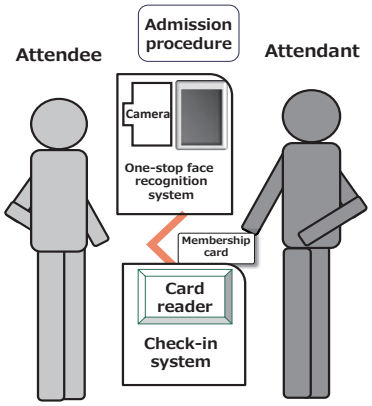


Fig. 3 One-stop face-recognition system

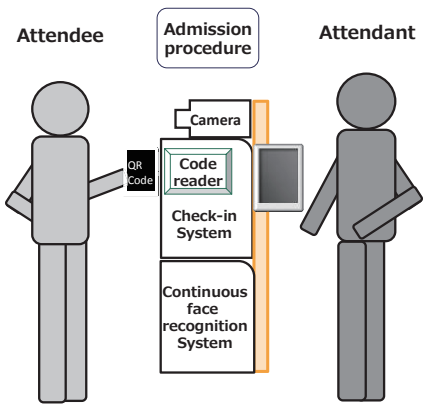


Fig. 4 Continuous face-recognition system

2) Managing eye contact

The proposed system uses a continuous-face-recognition system for accepting two photos of attendees successively taken with a single camera after an interval of 0.5 seconds to obtain facial photos with their eyes open. Few people spontaneously keep their eyes closed longer than 0.5 seconds because human blink duration is on average between 0.1 and 0.4 seconds [Bentivoglio 97]. Few people spontaneously blink twice in 0.5 seconds because human blink rate is between 7 and 17 per minute [Nosch 16]. It is possible to manage eye contact of attendees when we take the first photo at the same time of them scanning a QR code and then take the second photo after an interval of 0.5 seconds with a single camera.

3.2 Configuration of Proposed System

Figure 5 shows a configuration of the proposed system including event-attendee control platform and continuous-face-recognition system. The configuration is almost the same as that of the previous system [Okumura 17] except for a check-in doer and a QR code reader. While check-in doers of the previous system are attendants, those of the proposed system are attendees. While the previous system uses a card reader to scan membership cards, the proposed system uses a QR-code reader to scan tickets with QR codes. When attendees check in at a location that has the proposed system installed, they scan their QR-coded tickets. The attendee-management system provides the attendees the tickets in advance of the event day. Attendees can obtain tickets with QR codes with their smartphones. The ticket has the concert name, date and time, venue, QR code containing attendee's membership information, his/her name, seat number, registered photos, and so on.

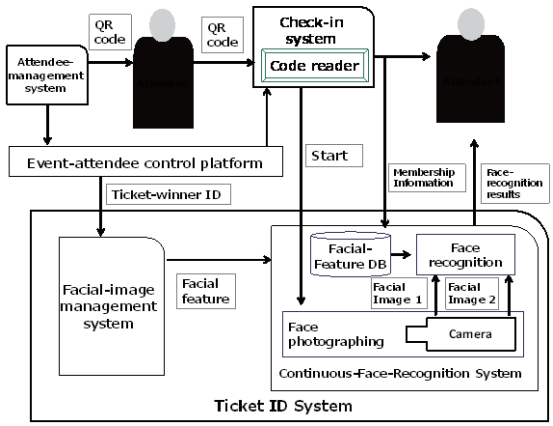


Fig. 5 Configuration of proposed system

3.3 Identity-Verification Procedure

An attendee's identity is verified with the procedure shown in Fig. 6. When attendees scan their QR codes, a check-in system performs ticket-winner check as well as showing the attendants the member information of the attendees, which is retrieved from the ticket-winner database with search keys of membership numbers obtained through a QR code reader. Scanning a QR code automatically activates continuous face recognition by taking two photos of the attendee after an interval of 0.5 seconds. When either photo is verified with the registered photo of the attendee, face recognition is successful. When attendees are ticket winners and face recognition is successful, the verification is successful. Otherwise, verification fails.

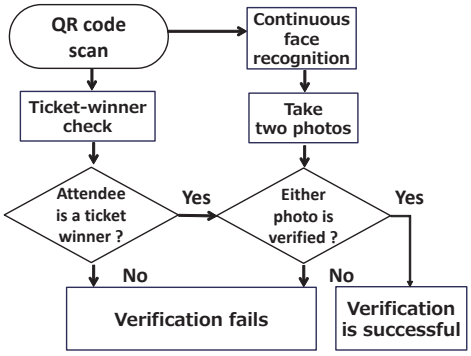


Fig. 6 Flowchart of verification

3.4 Operational Steps

The proposed system has the following operational steps from ticket application before the event day to admission on the event day:

Step 1: Ticket application is the same as that of the one-stop face-recognition system described in Section 2.1.

Step 2: Event organizers notify ticket winners, i.e., successful applicants that have been selected. They can obtain attendee's tickets including their QR codes and registered facial photos.

Step 3: At the check-in site on the event day, attendees scan their QR codes with the code reader according to attendant's instruction.

Step 4: Attendants can confirm that attendees are successful applicants who applied for the tickets.

Step 5: If identity is verified, the attendee is admitted entry. Otherwise, identity is verified by an attendant with direct visual inspection of the facial photo on the ticket.

4. Demonstration of Proposed System

4.1 Results

Six sets of the proposed system were used for a popular concert on November 17, 2018 in Tottori prefecture, Japan. Face recognition was carried out for 1,547 attendees. Face-recognition accuracy was 93%, and identity-verification time was 2.7 seconds on average in cases in which face recognition was not successful. No cases of attendees impersonating others were reported for the concert. The false reject rate (FRR) was 7% and the false accept rate (FAR) was 0%. There were two reasons for recognition failure: The first was that faces of incorrect attendees were detected when photos contained other attendees behind the correct attendee. The second was that the attendees had their faces covered with hair or items such as facemasks and mufflers. Failure was not observed due to the fact that attendees had their eyes closed or were not directly facing a camera.

4.2 Discussion

Table 1 compares the results of our previous and proposed systems. The average identity-verification time was 4.3 seconds shorter than that of the previous system because of changing check-in doers and improving face-recognition accuracy. The proposed system does not require handing over a membership

card between attendants and attendees. The face-recognition accuracy of the proposed system was 3% higher than that of the previous system. This resulted in decreasing the number of direct visual inspections, which increases identity-verification time.

There was a problem with face detection in that faces of incorrect attendees were detected when photos contained other people behind an attendee. This can be solved by choosing the face with the largest face area among all the detected faces. Face recognition failed when attendees had their faces covered with hair or items. There were no cases in which attendees had their eyes closed or were not directly facing a camera in the two successive photos, i.e., in the first photo and second photo taken after 0.5 seconds. Managing facial direction and eye contact of attendees worked as expected. It is difficult to solve the problem of when the faces of attendees are covered with hair or items because there would be no differences between the first and second photos on the attendee's covered faces. The attendee's cooperation is necessary for solving this problem.

Table 1 Results of our previous and proposed systems

	Previous system	Proposed system
Identity-verification time	7 seconds	2.7 seconds
Face-recognition accuracy	90%	93%
Check-in doer	Attendant	Attendee
Reasons for recognition failure	Attendees had their eyes closed.	
	Attendees were not directly facing camera.	
		Incorrect attendee's faces were detected.
	Attendees had their faces covered with their hair or items.	

5. Future Issues

Faces of incorrect attendees were detected when they stood behind a correct attendee. This can be solved by choosing the face with the largest face area among all the detected faces. If this improvement does now work, we are preparing a partitioning screen to be placed behind the correct attendee to prevent incorrect attendees from being photographed.

The largest obstacle remaining to improving face-recognition accuracy is that of covered faces. This problem could be solved with attendee's cooperation. We have been developing an identity-verification system using face recognition from selfies taken by attendees with their smartphone cameras [Okumura 18]. Self-photographing is regarded as helpful for securing clear facial photos because attendees can control intrinsic parameters such as their expressions, facial hair, and facial directions. We are planning to use of this system with the proposed system for solving the problem of covered faces.

The proposed system has been widely reported in the mass media. The system is highly regarded from reviews on the Internet [Hachima 18]. It was used to carry out face recognition for more than 100,000 attendees in 2018. Though no cases of attendees impersonating others were reported for any of these events, i.e., the FAR was 0%, the FAR should be more carefully examined from the view-point of preventing impersonation. It is necessary to evaluate the robustness against impersonation with pseudo attack tests. These tests should include disguise and lookalike tests. A disguise test makes people's facial appearances as similar to each other as possible by using facial paraphernalia such as facial hair, glasses, and makeup. A lookalike test is conducted for those, such as twins or similar looking siblings, with similar facial features. A disguise test will reveal considerable disguise methods and help in creating operational manuals for venue attendants to detect these methods. A lookalike test will disclose the technical limitations of current face-recognition methods and help in establishing next-generation technology.

6. Conclusion

We proposed an identity-verification system for attendees of large-scale events using continuous face recognition improved by managing check-in behavior of the attendees. The proposed system could secure facial photos of attendees directly facing a camera by leading them to scan their QR codes on a QR-code reader placed close to the camera just before executing face recognition. The system took two photos of attendees with this one single camera after an interval of 0.5 seconds to obtain facial photos with their eyes open. The system achieved 93% face-recognition accuracy with an average identity-verification time of 2.7 seconds per person when it was used for verifying 1,547 attendees at a concert of a popular music singer. The system made it possible to complete identity verification with higher accuracy with shorter average time than the previous system. We plan to improve our system to further streamline the verification procedure.

References

[Chapple 16] Chapple,J.: Ticket resale? NO, says Japanese music business (Aug. 23, 2016), available from < https://www.iq-mag.net/2016/08/ticket-resale-no-says-japanese-live-business-resaleno/#.W_01Jk8Un3g>.

[JE 15] JE fandom: Johnny's Tracks Illegally Sold Tickets for Arashi's Japonism Tour (Oct. 25, 2015), available from <https://jnewseng.wordpress.com/2015/10/25/johnnys-tracks-illegally-sold-tickets-for-arashis-japonism-tour/>.

[Okumura 17] Okumura, A., etc.: Identity Verification of Ticket Holders at Large-scale Events Using Face Recognition, *Journal of Information Processing*, Vol. 25, pp. 448-458 (Jun. 2017)

[NEC 17] NEC: Face Recognition, available from < https://www.nec.com/en/global/solutions/safety/face_recognition/index.html>.

[Bentivoglio 97] Bentivoglio AR, etc.: Analysis of blink rate patterns in normal subjects, Mov Disord. 12(6) pp1028-1034, (Nov. 1997)

[Nosch 16] Nosch DS, etc,: Relationship between Corneal Sensation, Blinking, and Tear Film Quality, Optom Vis Sci. ,93(5) pp471-481, (May. 2016)

[Okumura 18] Okumura, A., etc: Identity Verification for Attendees of Large-scale Events Using Face Recognition of Selfies Taken with Smartphone Cameras, *Journal of Information Processing*, Vol. 26, pp. 779-788 (Nov. 2018)

[Hachima 18] Hachima: Concerts of Hikaru Utada were successfully operated, available from < http://blog.esuteru.com/archives/9218587.html>. (Nov.2018) (in Japanese)